

**MEDINFO 2017: PRECISION HEALTHCARE THROUGH
INFORMATICS**

Studies in Health Technology and Informatics

This book series was started in 1990 to promote research conducted under the auspices of the EC programmes' Advanced Informatics in Medicine (AIM) and Biomedical and Health Research (BHR) bioengineering branch. A driving aspect of international health informatics is that telecommunication technology, rehabilitative technology, intelligent home technology and many other components are moving together and form one integrated world of information and communication media. The series has been accepted by MEDLINE/PubMed, SciVerse Scopus, EMCare, Book Citation Index – Science and Thomson Reuters' Conference Proceedings Citation Index.

Series Editors:

B. Blobel, O. Bodenreider, E. Borycki, M. Braunstein, C. Bühler, J.P. Christensen, R. Cooper, R. Cornet, J. Dewart, O. Le Dour, P.C. Dykes, A. Famili, M. González-Sancho, E.J.S. Hovenga, J.W. Jutai, Z. Kolitsi, C.U. Lehmann, J. Mantas, V. Maojo, A. Moen, J.F.M. Molenbroek, G. de Moor, M.A. Musen, P.F. Niederer, C. Nøhr, A. Pedotti, O. Rienhoff, G. Riva, W. Rouse, K. Saranto, M.J. Scherer, S. Schürer, E.R. Siegel, C. Safran, N. Sarkar, T. Solomonides, E. Tam, J. Tenenbaum, B. Wiederhold and L.H.W. van der Woude

Volume 245

Recently published in this series

- Vol. 244. R. Engelbrecht, R. Balicer and M. Hercigonja-Szekeres (Eds.), The Practice of Patient Centered Care: Empowering and Engaging Patients in the Digital Era
- Vol. 243. R. Röhrig, A. Timmer, H. Binder and U. Sax (Eds.), German Medical Data Sciences: Visions and Bridges – Proceedings of the 62nd Annual Meeting of the German Association of Medical Informatics, Biometry and Epidemiology (gmds e.V.) 2017 in Oldenburg (Oldenburg) – GMDS 2017
- Vol. 242. P. Cudd and L. de Witte (Eds.), Harnessing the Power of Technology to Improve Lives
- Vol. 241. C. Nøhr, C.E. Kuziemyky and Z.S.-Y. Wong (Eds.), Context Sensitive Health Informatics: Redesigning Healthcare Work
- Vol. 240. R.A. Logan and E.R. Siegel (Eds.), Health Literacy – New Directions in Research, Theory and Practice
- Vol. 239. A. Ryan, L.K. Schaper and S. Whetton (Eds.), Integrating and Connecting Care – Selected Papers from the 25th Australian National Health Informatics Conference (HIC 2017)
- Vol. 238. J. Mantas, A. Hasman, P. Gallos and M.S. Househ (Eds.), Informatics Empowers Healthcare Transformation
- Vol. 237. B. Blobel and W. Goossen (Eds.), pHealth 2017 – Proceedings of the 14th International Conference on Wearable Micro and Nano Technologies for Personalized Health, 14–16 May 2017, Eindhoven, The Netherlands
- Vol. 236. D. Hayn and G. Schreier (Eds.), Health Informatics Meets eHealth – Digital Insight – Information-Driven Health & Care – Proceedings of the 11th eHealth2017 Conference
- Vol. 235. R. Randell, R. Cornet, C. McCowan, N. Peek and P.J. Scott (Eds.), Informatics for Health: Connected Citizen-Led Wellness and Population Health
- Vol. 234. F. Lau, J. Bartle-Clar, G. Bliss, E. Borycki, K. Courtney and A. Kuo (Eds.), Building Capacity for Health Informatics in the Future
- Vol. 233. A.M. Kanstrup, A. Bygholm, P. Bertelsen and C. Nøhr (Eds.), Participatory Design & Health Information Technology

ISSN 0926-9630 (print)

ISSN 1879-8365 (online)

MEDINFO 2017: Precision Healthcare through Informatics

Proceedings of the 16th World Congress on Medical and Health Informatics

Edited by

Adi V. Gundlapalli

University of Utah/VA Salt Lake City Health Care System

Marie-Christine Jaulent

DR Inserm

and

Dongsheng Zhao

Academy of Military Medical Science of China

IOS
Press

Amsterdam • Berlin • Tokyo • Washington, DC

© 2017 International Medical Informatics Association (IMIA) and IOS Press.

This book is published online with Open Access and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).

ISBN 978-1-61499-829-7 (print)

ISBN 978-1-61499-830-3 (online)

Library of Congress Control Number: 2017963336

Publisher

IOS Press BV

Nieuwe Hemweg 6B

1013 BG Amsterdam

Netherlands

fax: +31 20 687 0019

e-mail: order@iospress.nl

For book sales in the USA and Canada:

IOS Press, Inc.

6751 Tepper Drive

Clifton, VA 20124

USA

Tel.: +1 703 830 6300

Fax: +1 703 830 2300

sales@iospress.com

LEGAL NOTICE

The publisher is not responsible for the use which might be made of the following information.

PRINTED IN THE NETHERLANDS

From the MedInfo2017 Editorial Committee: Templates, Submissions, and Camera-Ready Proceedings

Adi V. Gundlapalli^{a,b}

^a VA Salt Lake City Health Care System, Salt Lake City, UT, USA

^b Departments of Internal Medicine and Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA

Welcome to MedInfo2017

On behalf of the Editorial Committee (EC) for MedInfo2017, please accept our warm welcome to Hangzhou to celebrate the art and science of informatics in a truly breath-taking venue and with colleagues from around the world.

Learning from the experience of prior chairs

Learning from prior work and the experience of those who walked the path before is invaluable when embarking on a new venture. The first emails and calls after receiving word of my nomination to be EC chair were to prior chairs Christoph U. Lehmann (MedInfo2013) and Indra Neil Sarkar (MedInfo2015). While assuring me that this was a task that could be done (as they had so wonderfully demonstrated!), they did tell me that it was a huge effort that would take time, resources, and patience. It would be worth all the effort, they added. I am grateful to them for their guidance and gracious sharing of their experiences and lessons learned.

The journey

Starting with assisting the Scientific Program Committee (SPC) to bring out the call for papers, to witnessing the submissions start to come in as a trickle and end with an avalanche, to planning and executing the copy editing and formatting of the accepted papers and posters, the journey has been an interesting one to say the least.

First, let me start by saying what a great opportunity it has been to serve the international informatics community through MedInfo2017. It is an extraordinary honor to be able to chair the EC and I am grateful to the leadership of MedInfo2017 and the International Medical Informatics Association (IMIA) for their trust in me.

Second, I must acknowledge that it has truly been a great experience, as my predecessors had indicated it would be. Conceptualizing and executing the process of copy editing and formatting of nearly 400 five-page papers and one-page posters is a daunting task. I was ably assisted by two senior associate editors who are two amazing informatics students and rising stars in the field: Shauna Overgaard (University of Minnesota) and Kate Fultz Hollis (Oregon Health & Science University). Brainstorming, planning, and finally executing the work of the EC became a pleasure and I looked forward to our weekly conference calls to assess progress and resolve issues.

Third, working with Shauna and Kate, we recruited 41 outstanding associate editors (AEs) who then set out to copy edit and check the formatting of the accepted submissions. We set-up a multi-phase, cross-over plan wherein one AE would perform the editing and formatting and later another AE would review the work and revise as needed. The third phase involved

the senior AEs reviewing the submissions and finally, I personally reviewed each and every submission and then finalized the documents for the proceedings.

The editorial task is not trivial; the authors represent all continents (except Antarctica), and hail from varied backgrounds and professions, stages of training and career, and thereby epitomize a wide-range of experience in our field. Imagine a multifaceted channel in which technical research findings are communicated, by diverse researchers, often through a language barrier, to a diverse audience, who have been instructed to write academically in the official language of MedInfo (English). Now, add to that the requirement to adhere to a strict template (that is of paramount importance for uniformity and layout for the publishers). Indeed, this sets the stage for our work. But, be certain that it is undeniably humbling to experience the breadth, depth, and quality of the work presented by authors at MedInfo2017 and the proficiency with which their findings are communicated to an international audience.

The credit for preparing the submissions for the official conference proceedings goes to the AE team, the senior AEs, and the Editorial Committee. Even through a process where at least four individuals have reviewed a submission, it is possible that we may have missed a word here or formatting there. We offer our apologies for any inadvertent errors in the final proceedings.

The final stretch

It truly takes a village to complete such a large task. We acknowledge and thank the publishers, IOS Press, for their work on the proceedings. Paul Weij has been a great ally in this venture: offering advice and guidance on the technical aspects and being patient with us while we asked many questions and ultimately delivered the final files. Our thanks to the entire IOS Press team for their professional work and service with a smile. The VP of MedInfo2017, Kaija Saranto, has been an ardent supporter from the beginning. The SPC co-chairs, Elizabeth Borycki and Niels Peek, have been invaluable in their support and gracious with their advice in helping the EC complete its work. A special thanks and shout-out to Elaine Huesing, CEO, IMIA for her help in counseling and patiently guiding me through the work of the EC. Our thanks to Yongqin Huang (Chair), Emmett Huang (Vice-Chair), and the entire local organizing committee for their guidance in finalizing the proceedings. Our thanks to the two EC members, Dongsheng Zhao (China) and Marie-Christine Jaulent (France) for their involvement and culturally appropriate advice. And finally, listed in Table 1 are the AEs that spent countless hours reviewing, formatting, and copy-editing the nearly 400 papers and posters that make up the official MedInfo2017 conference proceedings.

Table 1 - Our dynamic and enthusiastic Assistant Editors for MedInfo2017

Assistant Editor (AE)	Institution
Abu Saleh Mohammad Mosa	University of Missouri School of Medicine, USA
Bryan Steitz	Vanderbilt University, USA
Celena B. Peters	VA Salt Lake City Health Care System and University of Utah School of Medicine, USA
Chen Liang	Louisiana Tech University, USA
Danny T. Y. Wu	University of Cincinnati, USA
David Chartash	Indiana University School of Medicine, USA
Doug Redd	George Washington University, USA
Andreas Triantafyllidis	Centre for Research and Technology Hellas (CERTH), Greece
Florence Femi Odekunle	Rutgers, The State University of New Jersey, USA
Elizabeth A. Lindemann	University of Minnesota, USA
Feichen Shen	Mayo Clinic, USA
Grace Gao	University of Minnesota and Robert Wood Johnson Foundation Future of Nursing Scholar Program, USA
Gregory W Hruby	Mount Sinai Medical Center, USA
Hyunggu Jung	University of Washington, USA
Jacob P. VanHouten	Vanderbilt University Medical Center, USA
Jenna Levenson Brager	Johns Hopkins University, USA
Jennifer Staley	Kent State University, USA
Kate Fultz Hollis (Senior AE)	Oregon Health & Science University, USA
Lacey Lewis	Veterans Health Administration, USA
Lisiane Pruinelli	University of Minnesota, USA
Lucy Lu Wang	University of Washington, USA
Marjorie E. Carter	University of Utah School of Medicine, USA
Mattias Georgsson	Blekinge Institute of Technology, Sweden
Maurine Tong	Columbia University Medical Center, USA
Preethi Srinivas	Regenstrief Institute Inc., USA
Rafeek Adeyemi Yusuf	University of Texas Health Science Center at Houston, USA
Ranyah Aldekhyyel	University of Minnesota, USA and King Saud University, Saudi Arabia
Rashmi Mishra	National Institute of Dental and Craniofacial Research, USA
Ryan A. McInerney	Queens University, Canada
Satish M. Mahajan	Veterans Health Administration, USA
Scott McGrath	University of Nebraska at Omaha, USA
Shauna M. Overgaard (Senior AE)	University of Minnesota, USA
Sripriya Rajamani	University of Minnesota, USA
Stacey Slager	VA Salt Lake City Health Care System and University of Utah School of Medicine, USA
Taya Irizarry	University of Pittsburgh, USA
Tiffany I. Leung	Maastricht University, The Netherlands
Tiffany J. Callahan	University of Colorado Denver, USA
Todd Lingren	University of Cincinnati, USA
Vignesh Subbian	University of Arizona, USA
Warren B P Pettey	VA Salt Lake City Health Care System and University of Utah School of Medicine, USA
Yanshan Wang	Mayo Clinic, USA
Yiye Zhang	Weill Cornell Medical College, Cornell University, USA
Zhe He	Florida State University, USA

As this editing experience carries prestige, is an enjoyable challenge, and is highly valuable for the CV, several members of the team have served as AEs for two and even three years. They have our gratitude and the satisfaction of having contributed to a worthy academic venture that will be read and cited for years to come. We have learned a few more lessons to add to the playbook created many years ago by our predecessors.

To the next MedInfo team, we wish you good luck and you know where to find us.

With warm regards and best wishes for a productive and successful MedInfo2017,

Adi V. Gundlapalli, MD, PhD, MS
Chair, Editorial Committee

From the Scientific Program Chairs

Elizabeth M. Borycki^a, Niels Peek^b

^a *School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada*

^b *Health eResearch Centre, University of Manchester, Manchester, United Kingdom*

The 16th World Congress of Medical and Health Informatics: Precision Healthcare through Informatics (MedInfo2017) was held in Hangzhou, China from August 21st to 25th, 2017, and celebrated the International Medical Informatics Association's (IMIA) 50th Anniversary. The conference was jointly hosted by IMIA and the Chinese Medical Informatics Association (CMIA).

China Medical Informatics Association

CMIA was established in 1980 and has provided over 37 years of leadership with its experts, scholars and managers leading and advancing medical informatics research and health information technologies in China and internationally. CMIA has over 6200 members, 115 directors, 18 regional branches, and 29 professional activities. CMIA draws its support from a broad and diverse group of organizations: the government, education, health and family planning, food and drug administration and traditional Chinese medicine ministries. We thank CMIA, its directors, professional groups and members for their support and their hosting MedInfo2017. We would also like to thank the Local Organizing Committee of MedInfo2017 for their hard work in organizing this outstanding conference.

MedInfo2017

The central theme of MedInfo2017 was "Precision Healthcare through Informatics", and alludes to the rapidly increasing opportunities to support customization of healthcare to the individual patient using the power of data and informatics. The scientific program was built up across five tracks: Connected and Digital Health; Human Data Science; Human, Organizational and Social Aspects; Knowledge Management and Quality, and Safety and Patient Outcomes.

We received 680 submissions (papers; posters; tutorials; panels; and workshops) from over 50 countries from all IMIA regions. The submissions spanned sub-disciplines in biomedical and health informatics such as clinical informatics, nursing informatics, consumer health informatics, public health informatics, human factors in health care, bioinformatics, translational informatics, quality and safety and research at the intersection of biomedical and health informatics, and precision medicine. After a thorough review process we accepted 249 papers, 168 posters, 18 tutorials, 23 panels, and 25 workshops. Each submission was reviewed independently by at least three reviewers, and subsequently assessed by one of the Scientific Program Committee (SPC) members and one of the SPC Track Chairs. Thus we ensured that all our decisions were well-informed and helped to create a program that is balanced across topics and has strong representation from all IMIA regions. Paper and poster submissions are included in these proceedings.

The conference also featured four keynote lectures given by researchers who are renowned for their work worldwide. Depei Liu from China is a leading researcher in area of data sharing in population health and the population link to health, science and technology. Harry Hemingway is conducting research in the United Kingdom, focused on innovating health and health care improvement through the use of data from life long patient records. Sue Bakken from the United States is presenting on research in the area of precision self-management. Lastly, Chris Seebregts leads several projects in the area of open medical health records and health information exchange focusing on how these technologies can be used to support public health system and reduce the burden of disease.

We would like to thank the Scientific Program Track Chairs, members, and the reviewers for their valuable contributions in creating the MedInfo2017 scientific program. We would also like to thank all contributors to MedInfo2017 without whom we would not be able to advance the science, education and practice of biomedical and health informatics globally.

International Medical Informatics Association

The International Medical Informatics Association (IMIA) celebrated its 50th Anniversary with MedInfo2017 in Hangzhou, China. IMIA has brought together academics, researchers and practitioners of biomedical and health informatics from around the world. It has led to the creation of a discipline and a field of study and education that is present in every country around the world. IMIA's global reach is significant and its leadership, through its members has led to the modernization of health care through information technology and to the support of citizens and health professionals use of technology throughout the world. Congratulations IMIA on your 50th Anniversary and we look forward to the next 50 years moving forward!

Sincerely,

Elizabeth Borycki, RN, PhD and Niels Peek, MSc, PhD
Co-Chairs, MedInfo2017 Scientific Program Committee

This page intentionally left blank

The International Medical Informatics Association (IMIA) Board 2017

Executive Committee

Hyeoun-Ae Park, President (2015 – 2017)
South Korea

Christoph Lehmann, President – Elect (2015 – 2017)
United States

Lincoln de Assis Moura Jr., Past President (2015 – 2017)
Brazil

Petter Hurlen, Secretary (2015 – 2018)
Norway

Sabine Koch, Treasurer (2015 – 2018)
Sweden

Vice Presidents of the IMIA Board

VP MedInfo (2015 – 2017)
Kaija Saranto
Finland

VP Membership (2012 – 2018)
Michio Kimura
Japan

VP Services (2016 – 2019)
Brigitte Seroussi
France

VP Special Affairs (2016 – 2019)
Elizabeth Borycki
Canada

VP WGs and SIGs (2016 – 2019)
Ying (Helen) Wu
China

Regional Vice Presidents to IMIA

VP for APAMI
Kyung-Hee Cho
South Korea

VP for EFMI
Anne Moen
Norway

VP for HELINA
Ghislain Kouematchoua Tchuitcheu
Cameroon

VP for IMIA-LAC
Amado Espinosa
Mexico

VP for MEAHI
Ramin Moghaddam
Iran

VP for North America
Andre Kushniruk
Canada

IMIA Liaison Officers

WHO Liaison Officer
Antoine Geissbuhler
Switzerland

IFIP Liaison Officer
Hiroshi Takeda
Japan

ISO Liaison Officer
Michio Kimura
Japan

This page intentionally left blank

Scientific Program Committee Co-Chairs

Elizabeth Borycki (Canada, Co-Chair)
Niels Peek (UK, Co-Chair)

Scientific Program Committee Members

Track I. Connected and Digital Health

Mowafa Househ, Saudi Arabia, Track Chair
Elizabeth Cummings, Australia
Rune Fensli, Norway
Luis Luque, Qatar
Michael Marscholke, Germany
Paula Otero, Argentina
Eustace Muteba Ayumba, DR Congo

Track II. Health Data Science

Fei Wang, US, Track Chair
Alex Kuo, Canada
Gregor Stiglic, Slovenia
Lisheng Xu, China
Nan Zhang, China
Goran Nenadic, UK
Evan Kontopantelis, UK
Allan Tucker, UK
Lucia Sacchi, Italy
Robert Moskovitch, Israel

Track III. Human, Organizational, and Social Aspects

Heather Strachan, UK, Track Chair
Sylvia Pelayo, France
Romaric Marcilly, France
NT Cheung, China, Hong Kong
Aviv Shachak, Canada
David Kaufman, US
Maria Häggglund, Sweden

Track IV. Knowledge Management

Patrick Weber, Switzerland, Track Chair
Ursula Hübner, Germany
Polun Chang, China, Taiwan
Marcos Didonet de Fabrio, Brazil
Max Topaz, Israel
Trevor Cohen, US

Track V. Quality and Safety, and Patient Outcomes

Patricia Dykes, US, Track Chair
Yang Gong, US
Rainer Röhrig, Germany
Yalini Senathirajah, US
Nicolette de Keizer, Netherlands
Daniel Rizzato Lede, Argentina
Paul Taylor, UK

Editorial Committee

Adi Gundlapalli (US, Chair)
Marie-Christine Jaulent (France)
Dongsheng Zhao (China)

VP MedInfo 2017

Kaija Saranto (Finland)

This page intentionally left blank

Local Organizing Committee Chair

Yongqin Huang

Local Organizing Committee Co-Chairs

Shuqian Luo
Baofen Ding
Haiyi Liu
Dongsheng Zhao
Ying Liang
Helen Wu
Meizhu Huang
Lieping Wu

Local Organizing Committee Vice Chairs

Serena Meng (Executive Vice Chair)
Yue Huang (Vice Chair)
Hua Li (Vice Chair)

Local Organizing Committee Members

Bright Chen
Michael Zhang
Maggie Liu
Lemon Liao
Sissi Shen
Phil Xiong
Cathy Zuo
Garry Yu

This page intentionally left blank

Contents

From the MedInfo2017 Editorial Committee: Templates, Submissions, and Camera-Ready Proceedings <i>Adi V. Gundlapalli</i>	v
From the Scientific Program Chairs <i>Elizabeth M. Borycki and Niels Peek</i>	vii
The International Medical Informatics Association (IMIA) Board 2017	ix
Scientific Program Committees	xi
Local Organizing Committees	xiii

Part 1

Papers

I. Connected and Digital Health

Multimodal e-Health Services for Smoking Cessation and Public Health: The SmokeFreeBrain Project Approach <i>Panagiotis D. Bamidis, Evangelos Paraskevopoulos, Evdokimos Konstantinidis, Dimitris Spachos and Antonis Billis</i>	5
Using Smartphones for Research Outside Clinical Settings: How Operating Systems, App Developers, and Users Determine Geolocation Data Quality in mHealth Studies <i>A.L. Beukenhorst, D.M. Schultz, J. McBeth, R. Lakshminarayana, J.C. Sergeant and W.G. Dixon</i>	10
The Future of Mobile Usability, Workflow and Safety Testing <i>Elizabeth Borycki, Yalini Senathirajah and Andre W. Kushniruk</i>	15
Integrated Care Processes Designed for the Future Healthcare System <i>Thomas Bürkle, Kerstin Denecke, Michael Lehmann, Erwin Zetz and Jürgen Holm</i>	20
Design and Evaluation on the Mobile Application of Transcutaneous Electrical Nerve Stimulation (TENS) <i>Ching-Lung Cheng, Li-Hui Lee and Yu-Ting Cheng</i>	25
In Search of a Digital Health Compass to Navigate the Health System <i>Catherine Chronaki, Robert Stegwee and Anne Moen</i>	30
Electronic Health Records in the Cloud: Improving Primary Health Care Delivery in South Africa <i>Liezel Cilliers and Graham Wright</i>	35
Decision-Tree Model for Support of Health Policy Choices Based on Pneumococcal Conjugate Vaccine (PCV) Program Outcomes <i>Andre Cohen, Casimir A. Kulikowski, Timothy Elbert, Samantha Clark, Dagna Constenla and Anushua Sinha</i>	40
Enabling Patient Control of Personal Electronic Health Records Through Distributed Ledger Technology <i>James Cunningham and John Ainsworth</i>	45
Enhancing Children's Safety by Barcoding Implementation at Breast Milk Feeding <i>Mariana Daus, Thelma Maydana, Daniel Rizzato Lede and Daniel Luna</i>	49
A Mobile System for Music Anamnesis and Receptive Music Therapy in the Personal Home <i>Kerstin Denecke</i>	54
Use of Flowchart for Automation of Clinical Protocols in mHealth <i>Karine Nóra Dias, Daniel Welfer, Marcos Cordeiro d'Ornellas, Carlos Jesus Pereira Haygert and Gustavo Nogara Dotto</i>	59

Tracking the Implementation of Electronic Medical Records in Dubai, United Arab Emirates, Using an Adoption Benchmarking Tool	64
<i>Osama El-Hassan, Amer Sharif, Mohammad Al Redha and Iain Blair</i>	
A Collaborative Evaluation Framework for Biometric Connected Devices in Healthcare	69
<i>Troskah Farnia, Marie Christine Jaulent, Guillaume Marchand and Mobin Yasini</i>	
Serious Games for Parkinson's Disease Fine Motor Skills Rehabilitation Using Natural Interfaces	74
<i>Antônio Augusto Foletto, Marcos Cordeiro d'Ornellas and Ana Lucia Cervi Prado</i>	
Patient Portal Adoption Rates: A Systematic Literature Review and Meta-Analysis	79
<i>Paolo Fraccaro, Markel Vigo, Panagiotis Balatsoukas, Iain E. Buchan, Niels Peek and Sabine N. van der Veer</i>	
E-HOSPITAL – A Digital Workbench for Hospital Operations and Services Planning Using Information Technology and Algebraic Languages	84
<i>Daniel Gartner and Rema Padman</i>	
Consumer Health Informatics Aspects of Direct-to-Consumer Personal Genomic Testing	89
<i>Kathleen Gray, Remya Stephen, Bronwyn Terrill, Brenda Wilson, Anna Middleton, Rigan Tytherleigh, Erin Turbitt, Clara Gaff, Jacqueline Savard, Chriselle Hickerton, Ainsley Newson and Sylvia Metcalfe</i>	
Designing a Mobile Health Application Prototype for the Management of Interstitial Cystitis/Painful Bladder Syndrome	94
<i>Janessa Griffith</i>	
cHRV Uncovering Daily Stress Dynamics Using Bio-Signal from Consumer Wearables	98
<i>Tian Hao, Henry Chang, Marion Ball, Kun Lin and Xinxin Zhu</i>	
Who Is Your Doctor? Analysis of Patient-Reported and EHR-Imputed Primary Care Physician	103
<i>Shan He, Sue Dintelman, Jake Sangster, Darren K. Mann, Thierry Guillermin and Sidney N. Thornton</i>	
Using Cloud-Based Physical Activity Data from Google Fit and Apple Healthkit to Expand Recording of Physical Activity Data in a Population Study	108
<i>André Henriksen, Laila Arnesdatter Hopstock, Gunnar Hartvigsen and Sameline Grimsgaard</i>	
Making Sense of Patient-Generated Health Data for Interpretable Patient-Centered Care: The Transition from “More” to “Better”	113
<i>Pei-Yun Sabrina Hsueh, Sanjoy Dey, Subhro Das and Thomas Wetter</i>	
Extraction and Quantification Clusters of Three-Dimensional Lorenz Plots	118
<i>Min Hu, Weixing Han, Jun Liu, Chengfan Jang and Suxia Wang</i>	
E-Health Literacy and Health Information Seeking Behavior Among University Students in Bangladesh	122
<i>Md. Mohaimenul Islam, Musa Touray, Hsuan-Chia Yang, Tahmina Nasrin Poly, Phung-Anh Nguyen, Yu-Chuan (Jack) Li and Shabbir Syed Abdul</i>	
Comparing Cancer Information Needs for Consumers in the US and China	126
<i>Zongcheng Ji, Yaoyun Zhang, Jun Xu, Xiaoling Chen, Yonghui Wu and Hua Xu</i>	
Acceptance and Use of eHealth/mHealth Applications for Self-Management Among Cancer Survivors	131
<i>Yun Jiang, Brady T. West, Debra L. Barton and Marcelline R. Harris</i>	
Promotion of Adequate Exercise for Chronic Disorders' Elderly Through Paced Music	136
<i>Gi-Ryong Kang, Il Kon Kim, Woo Jin Kim and Do-Youn Lee</i>	
Development of Parkinson Patient Generated Data Collection Platform Using FHIR and IoT Devices	141
<i>Dae-Young Kim, Sun-ho Hwang, Min-Gyu Kim, Joon-Hyun Song, Sin-Woong Lee and Il Kon Kim</i>	
Developing a Mobile Wellness Management System for Healthy Lifestyle by Analyzing Daily Living Activities	146
<i>Chae-Gyun Lim, Zae Myung Kim and Ho-Jin Choi</i>	
Guidance Through Use: Value as a Pathfinder in e-Health Services Implementation	151
<i>Bertil Lindenfalk and Vivian Vimarlund</i>	
Developing “Information Assistant”: A Smartphone Application to Meet the Personalized Information Needs of Women with Breast Cancer	156
<i>Ying Liu, Zhaohui Geng, Fulei Wu and Changrong Yuan</i>	
The Study of Smartphone Usage Competency Assessment and Training for the Elderly	161
<i>Sheng-Chieh Lu, Tzu-Ning Wen and Po-Lun Chang</i>	

Health Smart Homes: New Challenges <i>Anthony J. Maeder and Patricia A.H. Williams</i>	166
Dynamic Creation of Patient Summaries: A CDA and IHE XDS Based Approach for Regional EHRs <i>Patrick Mangesius, Samrend Saboor, Thomas Healy and Thomas Schabetsberger</i>	170
Analysis of HL7 EHRS Functional Model and Suggested Applications in China <i>Haibin Meng, Huajian Mao, Chenyang Chi and Dongsheng Zhao</i>	174
Are Health Literacy and eHealth Literacy the Same or Different? <i>Helen Monkman, Andre W. Kushniruk, Jeff Barnett, Elizabeth M. Borycki, Leigh E. Greiner and Debra Sheets</i>	178
An Exploratory Analysis of Game Telemetry from a Pediatric mHealth Intervention <i>Rema Padman, Disha Gupta, Bhargav Sri Prakash, Chelladurai Krishnan and K. Panchatcharam</i>	183
Acceptance of Using an Ecosystem of Mobile Apps for Use in Diabetes Clinic for Self-Management of Gestational Diabetes Mellitus <i>Sarita Pais, Dave Parry, Krassie Petrova and Janet Rowan</i>	188
Establishing Therapeutic Alliance in Mental Health Care via Cooperative Documentation <i>Anja Perlich, Julia von Thienen and Christoph Meinel</i>	193
Integrating Social Networks and Remote Patient Monitoring Systems to Disseminate Notifications <i>Hugo A. Ribeiro, Eliseu Germano, Sergio T. Carvalho and Eduardo S. Albuquerque</i>	198
Trust Model for Protection of Personal Health Data in a Global Environment <i>Pekka Ruotsalainen and Bernd Blobel</i>	202
PubMedReco: A Real-Time Recommender System for PubMed Citations <i>Hamman W. Samuel and Osmar R. Zaiane</i>	207
Looking for the Best WOW: Understanding the Nurses' Needs <i>Bibiana Schachner, Zulma González, Rodrigo Cano, Daniel Luna and Sonia Benítez</i>	212
Lessons Learned After Redesigning a Personal Health Record <i>Maria I. Smith, Gabriela Garcia, Mariana Simon, Lucila Bruchanski, Julia Frangella, Janine A. Sommer, Maria V. Giussi and Daniel R. Luna</i>	216
Mobile Application for Pregnant Women: What Do Mothers Say? <i>Janine Sommer, Mariana Daus, Maria Smith and Daniel Luna</i>	221
Operational Data Model Conversion to ResearchKit <i>Iñaki Soto-Rey, Sophia Gefner and Martin Dugas</i>	225
A Secure Architecture to Provide a Medical Emergency Dataset for Patients in Germany and Abroad <i>Michael Storck, Jan Wohlmann, Sarah Krudwig, Alexander Vogel, Judith Born, Thomas Weber, Martin Dugas and Christian Juhra</i>	230
An Architecture for Semantically Interoperable Electronic Health Records <i>André Toffanello, Ricardo Gonçalves, Adriana Kitajima, Ricardo Puttini and Atualpa Aguiar</i>	235
A Healthy Lifestyle Intervention Application <i>Trond Tufte and Ankica Babic</i>	240
A Case-Based Study with Radiologists Performing Diagnosis Tasks in Virtual Reality <i>José Eduardo Venson, Jean Carlo Albiero Berni, Carlos Edmilson da Silva Maia, Ana Maria Marques da Silva, Marcos Cordeiro d'Ornellas and Anderson Maciel</i>	244
Preferred Features of E-Mental Health Programs for Prevention of Major Depression in Male Workers: Results from a Canadian National Survey <i>Jian Li Wang and Kendall Ho</i>	249
Development of a Smartphone Application to Monitor Pediatric Patient-Reported Outcomes <i>Jingting Wang, Nengliang (Aaron) Yao, Yanyan Liu, Zhaohui Geng, Yuanyuan Wang, Nanpin Shen, Xiaoyan Zhang, Min Shen and Changrong Yuan</i>	253
I Got 99 Problems, and eHealth Is One <i>Sofie Wass and Vivian Vimarlund</i>	258

iT2DM Project: A Framework for Secondary Use of EHR Data for High-Throughput Phenotyping in Diabetes <i>Yufang Wei, Yujuan Shang, Jie Wang, Defu He, Shengrong Zhou, Weixun Zhang, Lili Shi, Yalan Chen, Kui Jiang, Huiqun Wu and Jiancheng Dong</i>	263
Development of Home-Based Sleep Monitoring System for Obstructive Sleep Apnea <i>Peirong Wu, Guan-Ting Chen, Yanyan Cui, Jin-Wei Li, Terry B.J. Kuo and Polun Chang</i>	268
II. Health Data Science	
Screening for Behavioral Risks: A Precision Healthcare Driven Approach for Chronic Pain Evaluation in Pediatric Specialty Care <i>Vibha Anand, Andrew S. Zeff and Steven J. Spalding</i>	275
Is There a Priority Shift in Mental Health Clinical Trials? <i>Vibha Anand, Soumya Ghosh and Amit Anand</i>	280
MedEx/J: A One-Scan Simple and Fast NLP Tool for Japanese Clinical Texts <i>Eiji Aramaki, Ken Yano and Shoko Wakamiya</i>	285
Locally Adaptive Operators for Red Lesions Detection in Eye Fundus Images <i>Luiza Dri Bagesteiro, Daniel Welfer, Marcos Cordeiro d'Ornellas, Juliano F. Kazienko, Carlos Jesus Pereira Haygert and Gustavo Nogara Dotto</i>	289
Semi-Automatic Mark-Up and UMLS Annotation of Clinical Guidelines <i>Matthias Becker and Britta Böckmann</i>	294
Inter-Annotator Agreement and the Upper Limit on Machine Performance: Evidence from Biomedical Natural Language Processing <i>Mayla Boguslav and Kevin Bretonnel Cohen</i>	298
Sharing Health Big Data for Research – A Design by Use Cases: The INSHARE Platform Approach <i>Guillaume Bouzillé, Richard Westerlynck, Gautier Defossez, Dallel Bouslimi, Sahar Bayat, Christine Riou, Yann Busnel, Clara Le Guillou, Jean-Michel Cauvin, Christian Jacquelinet, Patrick Pladys, Emmanuel Oger, Eric Stindel, Pierre Ingrand, Gouenou Coatrieux and Marc Cuggia</i>	303
Developing Visual Thinking in the Electronic Health Record <i>Andrew D. Boyd, Christine D. Young, Margret Amatayakul, Michael G. Dieter and Lawrence M. Pawola</i>	308
What Information Does Your EHR Contain? Automatic Generation of a Clinical Metadata Warehouse (CMDW) to Support Identification and Data Access Within Distributed Clinical Research Networks <i>Philipp Bruland, Justin Doods, Michael Storck and Martin Dugas</i>	313
Automatic Identification of Glaucoma Using Deep Learning Methods <i>Allan Cerentini, Daniel Welfer, Marcos Cordeiro d'Ornellas, Carlos Jesus Pereira Haygert and Gustavo Nogara Dotto</i>	318
Mining Adverse Drug Reactions in Social Media with Named Entity Recognition and Semantic Methods <i>Xiaoyi Chen, Myrtille Deldossi, Rim Aboukhamis, Carole Faviez, Badisse Dahamna, Pierre Karapetiantz, Armelle Guenegou-Arnoux, Yannick Girardeau, Sylvie Guillemin-Lanne, Agnès Lillo-Le-Louët, Nathalie Texier, Anita Burgun and Sandrine Katsahian</i>	322
Predicting Adverse Outcomes in Heart Failure Patients Using Different Frailty Status Measures <i>Yan Cheng, Yijun Shao, Charlene R. Weir, Rashmee U. Shah, Bruce E. Bray, Jennifer H. Garvin and Qing Zeng-Treitler</i>	327
A Data-Driven Decision-Support Tool for Population Health Policies <i>Michal Chorev, Lavi Shpigelman, Peter Bak, Avi Yaeli, Edwin Michael and Ya'ara Goldschmidt</i>	332
The Pluripotent Rendering of Clinical Data for Precision Medicine <i>Christopher G. Chute and Stanley M. Huff</i>	337
Classification of Clinical Research Study Eligibility Criteria to Support Multi-Stage Cohort Identification Using Clinical Data Repositories <i>James J. Cimino, William J. Lancaster and Mathew C. Wyatt</i>	341
Translational Morphosyntax: Distribution of Negation in Clinical Records and Biomedical Journal Articles <i>K. Bretonnel Cohen, Foster R. Goss, Pierre Zweigenbaum and Lawrence E. Hunter</i>	346

Extracting Sexual Trauma Mentions from Electronic Medical Notes Using Natural Language Processing <i>Guy Divita, Emily Brignone, Marjorie E. Carter, Ying Suo, Rebecca K. Blais, Matthew H. Samore, Jamison D. Fargo and Adi V. Gundlapalli</i>	351
General Symptom Extraction from VA Electronic Medical Notes <i>Guy Divita, Gang Luo, Le-Thuy T. Tran, T. Elizabeth Workman, Adi V. Gundlapalli and Matthew H. Samore</i>	356
An Integrated Surveillance System to Examine Testing, Services, and Outcomes for Sexually Transmitted Diseases <i>Brian E. Dixon, Guoyu Tao, Jane Wang, Wanzhu Tu, Sarah Hoover, Zuoyi Zhang, Teresa A. Batteiger and Janet N. Arno</i>	361
Development and Validation of Various Phenotyping Algorithms for Diabetes Mellitus Using Data from Electronic Health Records <i>Santiago Esteban, Manuel Rodríguez Tablado, Francisco Peper, Yamila S. Mahumud, Ricardo I. Ricci, Karin Kopitowski and Sergio Terrasa</i>	366
Detecting Signals of Interactions Between Warfarin and Dietary Supplements in Electronic Health Records <i>Yadan Fan, Terrence J. Adam, Reed McEwan, Serguei V. Pakhomov, Genevieve B. Melton and Rui Zhang</i>	370
From Bench to Bedside: A View on Bioinformatics Pipelines <i>Blanca Flores, Dirk Hose, Anja Seckinger, Petra Knaup and Matthias Ganzinger</i>	375
A Personalized Adviser-Interpreter System for Monitoring Diabetic Patients' Weight Loss <i>Azadeh Kamel Ghalibaf and Seyed Mahmood Tara</i>	379
An Infrared Thermal Images Database and a New Technique for Thyroid Nodules Analysis <i>José R. González, Charbel Damião and Aura Conci</i>	384
Automatically Identifying Topics of Consumer Health Questions in Chinese <i>Haihong Guo, Xu Na and Jiao Li</i>	388
Detecting Protected Health Information in Heterogeneous Clinical Notes <i>Aron Henriksson, Maria Kvist and Hercules Dalianis</i>	393
Acute Coronary Syndrome Risk Prediction Based on GRACE Risk Score <i>Danqing Hu, Zhengxing Huang, Tak-Ming Chan, Wei Dong, Xudong Lu and Huilong Duan</i>	398
Prediction and Factor Extraction of Drug Function by Analyzing Medical Records in Developing Countries <i>Min Hu, Yasunobu Nohara, Masafumi Nakamura and Naoki Nakashima</i>	403
Conditional Density Estimation of Tweet Location: A Feature-Dependent Approach <i>Hayate Iso, Shoko Wakamiya and Eiji Aramaki</i>	408
Analysis of the Existence of Patient Care Team Using Social Network Methods in Physician Communities from Healthcare Insurance Companies <i>Márcia Ito, Ana Paula Appel, Vagner Figueredo de Santana and Luis G. Moyano</i>	412
Measurement of Respiration Rate and Depth Through Difference in Temperature Between Skin Surface and Nostril by Using Thermal Image <i>Hieyong Jeong, Yutaka Matsuura and Yuko Ohno</i>	417
Visualizing Article Similarities via Sparsified Article Network and Map Projection for Systematic Reviews <i>Xiaonan Ji, Raghu Machiraju, Alan Ritter and Po-Yin Yen</i>	422
A Hybrid Method for ICD-10 Auto-Coding of Chinese Diagnoses <i>Zheng Jia, Weifeng Qin, Huilong Duan, Xudong Lv and Haomin Li</i>	427
The Impact of "Possible Patients" on Phenotyping Algorithms: Electronic Phenotype Algorithms Can Only Be Reproduced by Sharing Detailed Annotation Criteria <i>Rina Kagawa, Yoshimasa Kawazoe, Emiko Shinohara, Takeshi Imai and Kazuhiko Ohe</i>	432
Analysis of Historical Medical Phenomena Using Large N-Gram Corpora <i>Zdenko Kasáč and Stefan Schulz</i>	437
An Incremental Adoption Pathway for Developing Precision Medicine Based Healthcare Infrastructure for Underserved Settings <i>Suranga N. Kasthurirathne, Paul G. Biondich, Burke W. Mamlin, Theresa A. Cullen and Shaun J. Grannis</i>	442

Diagnostic Machine Learning Models for Acute Abdominal Pain: Towards an e-Learning Tool for Medical Students <i>Piyapong Khumrin, Anna Ryan, Terry Judd and Karin Verspoor</i>	447
Developing Methodologies to Find Abbreviated Laboratory Test Names in Narrative Clinical Documents by Generating High Quality Q-Grams <i>Kyungmo Kim and Jinwook Choi</i>	452
Characterising the Scope of Exposome Research: A Generalisable Approach <i>Philip Kiossoglou, Ann Borda, Kathleen Gray, Fernando Martin-Sanchez, Karin Verspoor and Guillermo Lopez-Campos</i>	457
Creation of the First French Database in Primary Care Using the ICPC2: Feasibility Study <i>V. Lacroix-Hugues, D. Darmon, C. Pradier and P. Staccini</i>	462
Conversion of National Health Insurance Service-National Sample Cohort (NHIS-NSC) Database into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) <i>Seng Chan You, Seongwon Lee, Soo-Yeon Cho, Hojun Park, Sungjae Jung, Jaehyeong Cho, Dukyong Yoon and Rae Woong Park</i>	467
Assessing Data Integration and Quality for the Evaluation of Point-of-Care Testing Across Rural and Remote Emergency Departments in Australia <i>Ling Li, Euan McCaughey, Juliana Iles-Mann, Andrew Sargeant, Maria R. Dahm, Virginia Mumford, Johanna I. Westbrook and Andrew Georgiou</i>	471
Using Machine Learning Models to Predict In-Hospital Mortality for ST-Elevation Myocardial Infarction Patients <i>Xiang Li, Haifeng Liu, Jingang Yang, Guotong Xie, Meilin Xu and Yuejin Yang</i>	476
Extracting Eligibility Criteria from the Narrative Text of Scientific Research Articles <i>Ching-Yun Lin, Der-Ming Liou and Mei-Lien Pan</i>	481
Assessing the Representation of Occupation Information in Free-Text Clinical Documents Across Multiple Sources <i>Elizabeth A. Lindemann, Elizabeth S. Chen, Sripriya Rajamani, Nivedha Manohar, Yan Wang and Genevieve B. Melton</i>	486
Precision Cohort Finding with Outcome-Driven Similarity Analytics: A Case Study of Patients with Atrial Fibrillation <i>Haifeng Liu, Xiang Li, Guotong Xie, Xin Du, Ping Zhang, Chengming Gu and Jingyi Hu</i>	491
Biomedical Informatics and the Digital Component of the Exposome <i>Guillermo Lopez-Campos, Mark Merolli and Fernando Martin-Sanchez</i>	496
Enhanced LexSynonym Acquisition for Effective UMLS Concept Mapping <i>Chris J. Lu, Destinee Tormey, Lynn McCreedy and Allen C. Browne</i>	501
A Validated Risk Model for 30-Day Readmission for Heart Failure <i>Satish M. Mahajan, Prabir Burman, Ana Newton and Paul A. Heidenreich</i>	506
CrowdMapping: A Crowdsourcing-Based Terminology Mapping Method for Medical Data Standardization <i>Huajian Mao, Chenyang Chi, Boyu Huang, Haibin Meng, Jinghui Yu and Dongsheng Zhao</i>	511
Case Report Form Reporter: A Key Component for the Integration of Electronic Medical Records and the Electronic Data Capture System <i>Yasushi Matsumura, Atsushi Hattori, Shiro Manabe, Daiyo Takahashi, Yuichiro Yamamoto, Taizo Murata, Akito Nakagawa, Naoki Mihara and Toshihiro Takeda</i>	516
Scholarly Information Extraction Is Going to Make a Quantum Leap with PubMed Central (PMC) [®] —But Moving from Abstracts to Full Texts Seems Harder than Expected <i>Franz Matthies and Udo Hahn</i>	521
Can Sonification Become a Useful Tool for Medical Data Representation? <i>George I. Mihalas, Minodora Andor, Anca Tudor and Sorin Paralescu</i>	526
Constructing a Gene-Drug-Adverse Reactions Network and Inferring Potential Gene-Adverse Reactions Associations Using a Text Mining Approach <i>MingShuang Sui and Lei Cui</i>	531
Establishment of New National Rare Disease (Nambyo) Registry and Registry Guidelines in Japan <i>Hiroshi Mizushima, Mai Tanabe, Yasutaka Sugamori, Yoko Sato and Hiromitsu Ogata</i>	536

Unsupervised Abbreviation Expansion in Clinical Narratives <i>Michel Oleynik, Markus Kreuzthaler and Stefan Schulz</i>	539
Contrast Mining for Pattern Discovery and Descriptive Analytics to Tailor Sub-Groups of Patients Using Big Data Solutions <i>Michael A. Phinney, Yan Zhuang, Sean Lander, Lincoln Sheets, Jerry C. Parker and Chi-Ren Shyu</i>	544
Detecting Adverse Drug Event Signals from a Clinical CaseBase <i>Weifeng Qin, Xian Zeng, Zheng Jia, Xiang Zheng, Huilong Duan, Xudong Lu and Haomin Li</i>	549
Extraction of Data from a Hospital Information System to Perform Process Mining <i>Ricardo Alfredo Quintano Neira, Gert-Jan de Vries, Jennifer Caffarel and Erin Stretton</i>	554
Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy <i>Manoj Raju, Venkatesh Pagidimarri, Ryan Barreto, Amrit Kadam, Vamsichandra Kasivajjala and Arun Aswath</i>	559
Envisioning Precision Healthcare Informatics: A Unified Framework <i>Arkalgud Ramaprasad, Nanda Kumar Bidare Sastry and Thant Syn</i>	564
A Virtual Earth Model of the Dementias in China <i>Hamish Robertson, Nick Nicholas, Jo Travaglia, Andrew Hayen and Andrew Georgiou</i>	569
Comparison of Regression Analysis and Transfer Function in Estimating the Parameters of Central Pulse Waves from Brachial Pulse Wave <i>Rui Chai, Li-sheng Xu, Yang Yao, Li-ling Hao and Lin Qi</i>	573
Longitudinal Changes in Risk Stratification for a Managed Population <i>Lincoln Sheets, Mihail Popescu, Kayson Lyttle, Soo-Yeon Cho and Jerry C. Parker</i>	578
Phenotypic Analysis of Clinical Narratives Using Human Phenotype Ontology <i>Feichen Shen, Liwei Wang and Hongfang Liu</i>	581
The Study of Pulse Condition of Traditional Chinese Medicine in Information Age <i>Cheng Shi, Yu Zhang, Qingyue Gong, Jinhai Zhou and Lisheng Xu</i>	586
Automated Evaluation System for Human Pupillary Behavior <i>Cleyton Rafael Gomes Silva, Cristhiane Gonçalves, Eduardo N.R. Camilo, Fabio Boaretti dos Santos, Joyce Siqueira, Eduardo Simões de Albuquerque, Fabrizio Alphonso Alves de Melo Nunes Soares, Leandro Luis Galdino de Oliveira and Ronaldo Martins da Costa</i>	589
Prescription Opioid Dependence in Western New York: Using Data Analytics to Find an Answer to the Opioid Epidemic <i>Shyamashree Sinha, Gale R. Burstein, Kenneth E. Leonard, Timothy F. Murphy and Peter L. Elkin</i>	594
Phenotyping and Visualizing Infusion-Related Reactions for Breast Cancer Patients <i>Deyu Sun, Gopal Sarda, Steven J. Skube, Anne H. Blaes, Saif Khairat, Genevieve B. Melton and Rui Zhang</i>	599
A Deep Learning-Based Method for Similar Patient Question Retrieval in Chinese <i>Guo Yu Tang, Yuan Ni, Guo Tong Xie, Xin Li Fan and Yan Ling Shi</i>	604
Automating the Identification of Patient Safety Incident Reports Using Multi-Label Classification <i>Ying Wang, Enrico Coiera, William Runciman and Farah Magrabi</i>	609
Mining Adverse Events of Dietary Supplements from Product Labels by Topic Modeling <i>Yefeng Wang, Divya R. Gunashekar, Terrence J. Adam and Rui Zhang</i>	614
An Early Infectious Disease Outbreak Detection Mechanism Based on Self-Recorded Data from People with Diabetes <i>Ashenafi Zebene Woldaregay, Eirik Årsand, Taxiarchis Botsis and Gunnar Hartvigsen</i>	619
Uncovering Hidden Topics in Hong Kong Clinical Research Through Hospital Authority Convention Publications <i>Zoie Shui-Yee Wong, Wai-Man Chan, Eliza Lai-Yi Wong, Patsy Yuen-Kwan Chau, Kwok-Leung Tsui and Hong Fung</i>	624
Aligned-Layer Text Search in Clinical Notes <i>Stephen Wu, Andrew Wen, Yanshan Wang, Sijia Liu and Hongfang Liu</i>	629
A Bayesian Network Model of Head and Neck Squamous Cell Carcinoma Incorporating Gene Expression Profiles <i>Shuyang Wu, Anthony Law and Mark E. Whipple</i>	634

Applying Risk Models on Patients with Unknown Predictor Values: An Incremental Learning Approach <i>Enliang Xu, Xiang Li, Jing Mei, Shiwan Zhao, Gang Hu, Eryu Xia, Haifeng Liu, Guotong Xie, Meilin Xu and Xuejun Li</i>	639
Semantic Relations in Compound Nouns: Perspectives from Inter-Annotator Agreement <i>Prabha Yadav, Elisabetta Jezek, Pierrette Bouillon, Tiffany J. Callahan, Michael Bada, Lawrence E. Hunter and K. Brettonel Cohen</i>	644
Graph Clustering System for Text-Based Records in a Clinical Pathway <i>Takanori Yamashita, Naoya Onimura, Hidehisa Soejima, Naoki Nakashima and Sachio Hirokawa</i>	649
Research on Ratio of Dosage of Drugs in Traditional Chinese Prescriptions by Data Mining <i>Xing-wen Yu, Qing-yue Gong, Kong-fa Hu, Wen-jing Mao and Wei-ming Zhang</i>	653
Retrofitting Concept Vector Representations of Medical Concepts to Improve Estimates of Semantic Similarity and Relatedness <i>Zhiguo Yu, Byron C. Wallace, Todd Johnson and Trevor Cohen</i>	657
“Hybrid Topics” – Facilitating the Interpretation of Topics Through the Addition of MeSH Descriptors to Bags of Words <i>Zhiguo Yu, Thang Nguyen, Ferdinand Dhombres, Todd Johnson and Olivier Bodenreider</i>	662
Identifying Biomarkers of Hepatocellular Carcinoma Based on Gene Co-Expression Network from High-Throughput Data <i>Ying Zhang, Zhiping Liu and Jing-song Li</i>	667
An Interactive Platform to Visualize Data-Driven Clinical Pathways for the Management of Multiple Chronic Conditions <i>Yiye Zhang and Rema Padman</i>	672

Part 2

III. Human, Organizational, and Social Aspects

Developing Digital Health and Welfare Services in an International Multidisciplinary Student Team <i>Otti Ahonen, Elina Rajalahti, Jonas Tana, Gun-Britt Lejongvist, Ulla-Mari Kinnunen and Kaija Saranto</i>	679
A User-Centered Glucose-Insulin Data Display for the Inpatient Setting <i>Michael E. Bowen, Umme Rumana, Ethan A. Killgore and Yang Gong</i>	684
Lack of Training for a Downtime Procedure: End User’s Perceptions of an Electronic Health Record Contingency Plan <i>Rodrigo Cano, Brenda Bejarano, Mercedes Vidal, Daniel Luna and Sonia Benítez</i>	689
Chile’s National Center for Health Information Systems: A Public-Private Partnership to Foster Health Care Information Interoperability <i>Daniel Capurro, Aisen Echeverry, Rosa Figueroa, Sergio Guiñez, Carla Taramasco, César Galindo, Angélica Avendaño, Alejandra García and Steffen Härtel</i>	693
A Combined Collaborative Information Behaviour (CIB) and Continuity of Care Framework for Modeling Complexity in Colorectal Cancer Screening Access <i>Caitlin Champion, Ewan Affleck, Gonzalo G. Alvarez and Craig Kuziemy</i>	696
Clinical Decision Support and Primary Care Acceptance of Genomic Medicine <i>Dian A. Chase, Sherry Baron and Joan S. Ash</i>	700
An Experimental Comparison of Quality Models for Health Data De-Identification <i>Johanna Eicher, Klaus A. Kuhn and Fabian Prasser</i>	704
Evaluation of the Association Between Different Patient Indexing Strategies and Effective Indexing in Health Centers of the Public System of the Autonomous City of Buenos Aires: An Exploratory Study <i>Santiago Esteban, Cecilia Palermo, Leandro Alassia, M. Victoria Giussi, Analía Baum and Fernán González Bernaldo de Quirós</i>	709
Design and Construction for Community Health Service Precision Fund Appropriation System Based on Performance Management <i>Xing Gao, Yao He and Hongpu Hu</i>	714

Ethical Issues in Implementing National-Level Health Data Warehouses in Developing Countries <i>Milka B. Gesicho, Troy D. Moon, Elizabeth Heitman and Martin C. Were</i>	718
Patients' Online Access to Electronic Health Records: Current Status and Experiences from the Implementation in Sweden <i>Maria Hägglund and Isabella Scandurra</i>	723
What Do They Mean by "Health Informatics"? Health Informations Posts Compared to Program Standards <i>Joseette F. Jones, Enming Zhang, Anand Kulanthaivel and Shilpa Katta</i>	728
Design of a User-Centered Voluntary Reporting System for Patient Safety Events <i>Hong Kang and Yang Gong</i>	733
OpenClinic GA Open Source Hospital Information System Enabled Universal Health Coverage Monitoring and Evaluation in Burundian Hospitals <i>Gustave Karara, Frank Verbeke, Evelyn Ndabaniwe, Etienne Mugisho and Marc Nyssen</i>	738
User Centred Design and Nosocomials in Surgical ICUs: A Mobile Application for Peer Monitoring and Training in Hand Hygiene <i>Nadish Kariyawasam, Ming Chao Wong and Paul Turner</i>	743
An Experimental Comparison of a Co-Design Visualizing Personal Drug Information and Patient Information Leaflets: Usability Aspects <i>Soudabeh Khodambashi, Dagrun Haugland, Anette Ellingsberg, Hanne Kottum, Janne Kutschera Sund and Øystein Nytrø</i>	748
Mutual Learning and Exchange of Health Informatics Experiences from Around the World – Evaluation of a Massive Open Online Course in eHealth <i>Sabine Koch and Maria Hägglund</i>	753
50th Anniversary International Medical Informatics Association (IMIA) History Working Group and Its Projects <i>Casimir A. Kulikowski, George Mihalas, Robert Greenes, Valerio Yacubsohn and Hyeoun-Ae Park</i>	758
Towards a Usability and Error "Safety Net": A Multi-Phased Multi-Method Approach to Ensuring System Usability and Safety <i>Andre Kushniruk, Yalini Senathirajah and Elizabeth Borycki</i>	763
The New HIT: Human Health Information Technology <i>Tiffany I. Leung, Mary K. Goldstein, Mark A. Musen, Ruth Cronkite, Jonathan H. Chen, Assaf Gottlieb and Eran Leitersdorf</i>	768
Global eHealth, Social Business and Citizen Engagement: A Natural Convergence? <i>Siaw-Teng Liaw, Alvin Marcelo, Padmanesan Narasimhan, Md. Mahfuz Ashraf and Pradeep Ray</i>	773
P-Care BPJS Acceptance Model in Primary Health Centers <i>Hosizah Markam</i>	778
The Australian Health Informatics Competencies Framework and Its Role in the Certified Health Informatician Australasia (CHIA) Program <i>Fernando Martin-Sanchez, David Rowlands, Louise Schaper and David Hansen</i>	783
Information and Communication Gaps in Intersectoral Healthcare Processes for Dementia Patients <i>Bianca Oppermann, Bettina Zippel-Schultz, Uta Augustin, Andrea Popa, Carsten Schultz and Reinhold Haux</i>	788
Equality Challenges in the Use of eHealth: Selected Results from a Danish Citizens Survey <i>Lone Stub Petersen and Pernille Bertelsen</i>	793
Analysis of the Updated Swedish Regulatory Framework of the Patient Accessible Electronic Health Record in Relation to Usage Experience <i>Isabella Scandurra, Maria Pettersson, Benny Eklund and Leif Lyttkens</i>	798
Green Turning Brown – Domain Engineering for Social and Health Services in Finland <i>Reima Suomi, Pirkko Nykänen, Tapio Vepsäläinen and Riina Hiltunen</i>	803
Toward Automatic Reporting of Infectious Diseases <i>Hsiao-Mei Tsao, Chi-Ming Chang, Jen-Hsiang Chuang, Ding-Ping Liu, Mei-Lien Pan and Da-Wei Wang</i>	808
Change Management Strategies: Transforming a Difficult Implementation into a Successful One <i>Maria Victoria Giusi, Analía Baum, Fernando Plazzotta, Pablo Muguerza and Fernán González Bernaldo de Quirós</i>	813

Evaluate the Usability of the Mobile Instant Messaging Software in the Elderly <i>Tzu-Ning Wen, Po-Liang Cheng and Po-Lun Chang</i>	818
Augmented Reality in Nursing: Designing a Framework for a Technology Assessment <i>Hanna Wüller, Marcus Garthaus and Hartmut Remmers</i>	823
Design a Learning-Oriented Fall Event Reporting System Based on Kirkpatrick Model <i>Sicheng Zhou, Hong Kang and Yang Gong</i>	828
IV. Knowledge Management	
Managing User Needs During the EHR Implementation in Buenos Aires City: The HelpDesk Role <i>L. Alassia, C. Palermo, F. Recondo, M. Giussi, A. Stieben, A. Baum and F. Gonzalez Bernaldo de Quiros</i>	835
Expressing Biomedical Ontologies in Natural Language for Expert Evaluation <i>Muhammad Amith, Frank J. Manion, Marcelline R. Harris, Yaoyun Zhang, Hua Xu and Cui Tao</i>	838
Eliciting the Intension of Drug Value Sets – Principles and Quality Assurance Applications <i>Nathan J. Bahr, Scott D. Nelson, Rainer Winnenburg and Olivier Bodenreider</i>	843
Comparison of Three English-to-Dutch Machine Translations of SNOMED CT Procedures <i>Ronald Cornet, Carly Hill and Nicolette de Keizer</i>	848
Trends in Fetal Medicine: A 10-Year Bibliometric Analysis of Prenatal Diagnosis <i>Ferdinand Dhombres and Olivier Bodenreider</i>	853
The Portal of Medical Data Models: Where Have We Been and Where Are We Going? <i>Sophia Gefner, Philipp Neuhaus, Julian Varghese, Philipp Bruland, Alexandra Meidt, Iñaki Soto-Rey, Michael Storck, Justin Doods and Martin Dugas</i>	858
Perceiving the Usefulness of the National Cancer Institute Metathesaurus for Enriching NCIt with Topological Patterns <i>Zhe He, Yan Chen and James Geller</i>	863
Shiny FHIR: An Integrated Framework Leveraging Shiny R and HL7 FHIR to Empower Standards-Based Clinical Data Applications <i>Na Hong, Naresh Prodduturi, Chen Wang and Guoqian Jiang</i>	868
Key Relation Extraction from Biomedical Publications <i>Lan Huang, Ye Wang, Leiguang Gong, Casimir Kulikowski and Tian Bai</i>	873
Knowledge Management in Health Technology SMEs <i>Juhamatti Huusko, Sirpa Kuusisto-Niemi and Kaija Saranto</i>	878
Effects of Implementing a Tree Model of Diagnosis into a Bayesian Diagnostic Inference System <i>Satoshi Iwai, Yoshimasa Kawazoe, Takeshi Imai and Kazuhiko Ohe</i>	882
A Consensus-Based Approach for Harmonizing the OHDSI Common Data Model with HL7 FHIR <i>Guoqian Jiang, Richard C. Kiefer, Deepak K. Sharma, Eric Prud'hommeaux and Harold R. Solbrig</i>	887
Text Data Mining of Aged Care Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes <i>Tao Jiang, Siyu Qian, David Hailey, Jun Ma and Ping Yu</i>	892
An Ontology-Based Approach to Estimate the Frequency of Rare Diseases in Narrative-Text Radiology Reports <i>Charles E. Kahn Jr.</i>	896
On Incentives for Open Access Publishing: A Survey at IMIA's Annual General Assembly During HEC2016 <i>Stefanie Kuballa, Mareike Schulze, Corinna Mielke, Olaf Gefeller and Reinhold Haux</i>	901
Peer Review of Clinical Information Models: A Web 2.0 Crowdsourced Approach <i>Heather Leslie and Silje Ljosland Bakke</i>	905
A Semi-Automatic Framework to Identify Abnormal States in EHR Narratives <i>Xiaojun Ma, Takeshi Imai, Emiko Shinohara, Ryota Sakurai, Kouji Kozaki and Kazuhiko Ohe</i>	910
Application of Ontology Technology in Health Statistic Data Analysis <i>Minjiang Guo, Hongpu Hu and Xingyun Lei</i>	915

Interoperability of Medication Classification Systems: Lessons Learned Mapping Established Pharmacologic Classes (EPCs) to SNOMED CT	920
<i>Scott D. Nelson, Jaqui Parker, Robert Lario, Rainer Winnenburg, Mark S. Erlbaum, Michael J. Lincoln and Olivier Bodenreider</i>	
Interoperability of Disease Concepts in Clinical and Research Ontologies: Contrasting Coverage and Structure in the Disease Ontology and SNOMED CT	925
<i>Satyajeet Raje and Olivier Bodenreider</i>	
From Patient Discharge Summaries to an Ontology for Psychiatry	930
<i>Marion Richard, Xavier Aimé, Marie-Christine Jaulent, Marie-Odile Krebs and Jean Charlet</i>	
Surveying a New Multi-Institution Clinical Data Research Network	935
<i>Marc Rosenman, Margaret Madden, Elissa Oh, Satyender Goel and Abel Kho</i>	
Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration	940
<i>Stefan Schulz, Jean-Marie Rodrigues, Alan Rector and Christopher G. Chute</i>	
Semantic Web Service Delivery in Healthcare Based on Functional and Non-Functional Properties	945
<i>Marco Schweitzer, Thilo Gorfer and Alexander Hörbst</i>	
An OMOP CDM-Based Relational Database of Clinical Research Eligibility Criteria	950
<i>Yuqi Si and Chunhua Weng</i>	
Characterizing Surgical Site Infection Signals in Clinical Notes	955
<i>Steven J. Skube, Zhen Hu, Elliot G. Arsoniadis, Gyorgy J. Simon, Elizabeth C. Wick, Clifford Y. Ko and Genevieve B. Melton</i>	
Formalizing Evidence Type Definitions for Drug-Drug Interaction Studies to Improve Evidence Base Curation	960
<i>Joseph Utecht, Mathias Brochhausen, John Judkins, Jodi Schneider and Richard D. Boyce</i>	
A Method for Constructing a New Extensible Nomenclature for Clinical Coding Practices in Sub-Saharan Africa	965
<i>Sven Van Laere, Marc Nyssen and Frank Verbeke</i>	
Content and Trends in Medical Informatics Publications over the Past Two Decades	968
<i>Liqin Wang, Maxim Topaz, Joseph M. Plasek and Li Zhou</i>	
Leveraging Event Reporting Through Knowledge Support: A Knowledge-Based Approach to Promoting Patient Fall Prevention	973
<i>Bin Yao, Hong Kang, Qi Miao, Sicheng Zhou, Chen Liang and Yang Gong</i>	
Summarizing an Ontology: A “Big Knowledge” Coverage Approach	978
<i>Ling Zheng, Yehoshua Perl, Gai Elhanan, Christopher Ochs, James Geller and Michael Halper</i>	
KQA: A Knowledge Quality Assessment Model for Clinical Decision Support Systems	983
<i>Seyedjamal Zolhavarieh and Dave Parry</i>	
V. Quality and Safety, and Patient Outcomes	
An Innovative Approach to Jointly Scheduling and Assigning a Consultation Time to Patients Arriving in the Emergency Department	989
<i>Faten Ajmi, Sarah Ben Othman, Hayfa Zgaya, Slim Hammadi and Jean-Marie Renard</i>	
Design of a Mobile Application for Transfusion Medicine	994
<i>M.A. Albornoz, S. Márquez, L. Rubin and D. Luna</i>	
Assessing Provider-Generated Free-Text Quality in EHR-Integrated Handoff Notes	999
<i>Elliot G. Arsoniadis, Steven J. Skube, Treva M. Bjerke, Bryan Jarabek and Genevieve B. Melton</i>	
Accessing Reliable Health Information on the Web: A Review of the HON Approach	1004
<i>Célia Boyer, Arnaud Gaudinat, Allan Hanbury, Ron D. Appel, Marion J. Ball, Michel Carpentier, Jan H. van Bommel, Jean-Paul Bergmans, Denis Hochstrasser, Donald Lindberg, Randolph Miller, Jean-Claude Peterschmitt, Charles Safran, Michèle Thonnet and Antoine Geissbühler</i>	
Arden Syntax Clinical Foundation Framework for Event Monitoring in Intensive Care Units: Report on a Pilot Study	1009
<i>Jeroen S. de Bruin, Julia Zeckl, Katharina Adlassnig, Alexander Blacky, Walter Koller, Andrea Rappelsberger and Klaus-Peter Adlassnig</i>	

Design of an Intelligent Nursing Clinical Pathway and Nursing Order Support System for Traditional Chinese Medicine <i>Bao-Fen Ding, Polun Chang, Ping Wang, Hai-Ting Li and Ming-Chuan Kuo</i>	1014
Predicting Length of Stay for Obstetric Patients via Electronic Medical Records <i>Cheng Gao, Abel N. Kho, Catherine Ivory, Sarah Osmundson, Bradley A. Malin and You Chen</i>	1019
Group-Based Trajectory Analysis for Long-Term Use of Warfarin Therapy in Atrial Fibrillation Patients <i>Shijing Guo, Xiang Li, Xin Du, Haifeng Liu and Guotong Xie</i>	1024
CTdBem – A New Protocol for Ultra Low Radiation Dose MDCT <i>Carlos Jesus Pereira Haygert, Marcos Cordeiro d'Ornellas, Daniel Welfer, Rubens Martins Bastos and Gustavo Nogara Dotto</i>	1029
Successful System, Incomplete Data: Caveats in Reusing Activity Data from Emergency Department Whiteboards <i>Morten Hertzum</i>	1033
Improving Patient Safety Through the Design and Development of a Computerized Provider Order Entry for Parenteral Nutrition Linked to a Barcode Medication Administration Record <i>Oscar I. Jáuregui, Lucila Bruchanski, Daniel A. Rizzato Lede, Carlos M. Otero and Daniel R. Luna</i>	1038
Use of EHR Data to Identify Factors Affecting the Time to Fall <i>Hyesil Jung and Hyeoun-Ae Park</i>	1043
Identifying and Synchronizing Health Information Technology (HIT) Events from FDA Medical Device Reports <i>Hong Kang, Frank Wang, Sicheng Zhou, Qi Miao and Yang Gong</i>	1048
Comparative Performance Analysis of Different Fingerprint Biometric Scanners for Patient Matching <i>Noah Kasiiti, Judy Wawira, Saptarshi Purkayastha and Martin C. Were</i>	1053
Development and Evaluation of a Blood Glucose Management System for Reducing the Delay in Measurement <i>Teramoto Kei, Okura Tsuyoshi, Koshitani Rie, Takahiro Murata, Yorifuji Yuko, Hotta Erika, Endo Yusuke, Ueki Masaru and Kondoh Hiroshi</i>	1058
Evaluation of a Decision Support System for Laboratory Service Patients <i>Georgy Kopanitsa</i>	1063
Effective Usability Engineering in Healthcare: A Vision of Usable and Safer Healthcare IT <i>Andre Kushniruk, Yalini Senathirajah and Elizabeth Borycki</i>	1066
Automated Classification of Multi-Labeled Patient Safety Reports: A Shift from Quantity to Quality Measure <i>Chen Liang and Yang Gong</i>	1070
Predicting Harm Scores from Patient Safety Event Reports <i>Chen Liang and Yang Gong</i>	1075
The Value of Data and the Value of Questions: Achieving Improvement in Hospital Pharmacy and Medication Safety <i>Valentina Lichtner and Tony Cornford</i>	1080
User-Centered Design Improves the Usability of Drug-Drug Interaction Alerts: A Validation Study in the Real Scenario <i>Daniel R. Luna, Daniel A. Rizzato Lede, Luciana Rubin, Carlos M. Otero, Juan M. Ortiz, Mónica G. García, Romina P. Rapisarda, Marcelo R. Risk and Fernán González Bernaldo de Quirós</i>	1085
Extracting Follow-Up Recommendations and Associated Anatomy from Radiology Reports <i>Thusitha Mabotuwana, Christopher S. Hall, Sandeep Dalal, Joel Tieder and Martin L. Gunn</i>	1090
BarCode Medication Administration in ICU: Learning from Our Nurses <i>Thelma Maydana, Liliana Giraldo, Zulma González, Bibiana Schachner, John Mayan, Daniel Luna and Sonia Benítez</i>	1095
Procedure for Reconstruction of a Predictive Score of Severe Deterioration in Inpatients <i>Hernán Navas and Elián Bourdin</i>	1099
Framework of Performance Measures for Health Information Exchange (HIE) <i>Okada Mihoko, Ishida Haku, Watanabe Sunao and Shiratori Yoshimune</i>	1103

A Comparison of Two Principal Systems for Monitoring of Technology-Induced Errors in Electronic Health Records <i>Sari Palojoki, Elizabeth M. Borycki, Andre W. Kushniruk and Kaija Saranto</i>	1108
Factors Affecting the Time to Occurrence of Hospital-Acquired Pressure Ulcers Using EHR Data <i>Seul Ki Park and Hyeoun-Ae Park</i>	1113
Precision of EMR Data: The Case for a Drug and Alcohol Service <i>Siyu Qian, Esther Munyisia, Ping Yu, David Hailey and David Reid</i>	1118
Adverse Drug Event Monitoring with Clinical and Laboratory Data Using Arden Syntax <i>Andrea Rappelsberger, Klaus-Peter Adlassnig, Jeroen S. de Bruin, Manuela Plössl, Jochen Schuler and Christina Hofer-Dückelmann</i>	1123
Usability Evaluation of an EHR's Clinical Notes Interface from the Perspective of Attending and Resident Physicians: An Exploratory Study <i>Rubina F. Rizvi, Jenna L. Marquard, Marcus A. Seywerd, Terrence J. Adam, Jed T. Elison, Gretchen M. Hultman, Kathleen A. Harder and Genevieve B. Melton</i>	1128
Strengthening Data Confidentiality and Integrity Protection in the Context of a Multi-Centric Information System Dedicated to Autism Spectrum Disorder <i>Mohamed Ben Said, Laurence Robel, Bernard Golse and Jean Philippe Jais</i>	1133
Mapping the Electronic Health Record: A Method to Study Display Fragmentation <i>Yalini Senathirajah, Jinglu Wang, Elizabeth Borycki and Andrek Kushniruk</i>	1138
Strengthen Cancer Surveillance in Sri Lanka by Implementing Cancer Registry Informatics to Enhance Cancer Registry Data Accuracy, Completeness, and Timeliness <i>Kamal Seneviratne, Sudath Samaraweera, Eshani Fernando, Suraj Perera and Asanga Hemarathna</i>	1143
Using Therapeutic Circles to Visualize Guideline-Based Therapeutic Recommendations for Patients with Multiple Chronic Conditions: A Case Study with GO-DSS on Hypertension, Type 2 Diabetes, and Dyslipidemia <i>Brigitte Séroussi, Alexandre Galopin, Mouhssin Gaouar, Suzanne Pereira and Jacques Bouaud</i>	1148
Development of a Service-Oriented Sharable Clinical Decision Support System Based on Ontology for Chronic Disease <i>Yong Shang, Yu Wang, Ling Gou, Chengkai Wu, Tianshu Zhou and Jing-song Li</i>	1153
The Paradox of Higher Charges for Lower-Risk Inpatient Admissions: When Healthier Patients Cost More <i>Lincoln Sheets, Kayson Lyttle, Lori L. Popejoy and Jerry C. Parker</i>	1158
A Design of a Surgical Site Verification System <i>Biyu Shen, Yan He and Haoyang Chen</i>	1163
Characterizing Restorative Dental Treatments of Sjögren's Syndrome Patients Using Electronic Dental Records Data <i>Zasim Siddiqui, Yue Wang, Payal Makkad and Thankam Thyvalikakath</i>	1166
Analysis of Clinical Variations in Asthma Care Documented in Electronic Health Records Between Staff and Resident Physicians <i>Sunghwan Sohn, Chung-Il Wi, Young J. Juhn and Hongfang Liu</i>	1170
Medical Effect of Venous Thromboembolism Prophylaxis Systems and Common Input Categories: Preliminary Findings from a Systematic Review <i>Julian Varghese, Maren Kleine, Sophia Isabella Gessner, Sarah Sandmann and Martin Dugas</i>	1175
CD-KES: An Ontology Based Knowledge Education System for Patients with Chronic Diseases and Its Constructing Approach <i>Chengkai Wu, Ling Gou, Yu Wang, Yong Shang, Yu Tian and Jing-song Li</i>	1180
Gathering Real World Evidence with Cluster Analysis for Clinical Decision Support <i>Eryu Xia, Haifeng Liu, Jing Li, Jing Mei, Xuejun Li, Enliang Xu, Xiang Li, Gang Hu, Guotong Xie and Meilin Xu</i>	1185
Context-Sensitive Clinical Alert Packages Written in Arden Syntax <i>Julia Zeckl, Katharina Adlassnig, Renate Fossler, Alexander Blacky, Jeroen S. de Bruin, Walter Koller, Andrea Rappelsberger and Klaus-Peter Adlassnig</i>	1190

- Validation of the Chinese Version of the Functional Oral Intake Scale (FOIS) Score in the Assessment of Acute Stroke Patients with Dysphagia 1195
Hongzhen Zhou, Yafang Zhu and Xiaomei Zhang

- The Impact of Censoring Drug Switching in Medication Adherence Measures of Chronic Single Ingredient Oral Drugs 1200
Vivienne J. Zhu, J. Marc Overhage, Qianli Ma and Patrick B. Ryan

Posters

I. Connected and Digital Health

- SMS Education for the Promotion of Diabetes Self-Management in Low & Middle Income Countries: A Randomized Controlled Trial in Egypt 1209
Haitham Abaza, Michael Marschollek and Mareike Schulze
- A Study on Saudi Diabetic Patients' Readiness to Use Mobile Health 1210
Hanan Alenazi, Mariam Alghamdi, Sarah Alradhi, Mowafa Househ and Nasriah Zakaria
- Saudi Diabetic Patients' Attitudes Towards Patient Portal Use and Their Perceived E-Health Literacy 1211
Rawan AlOthman, Nasriah Zakaria and Ahmed AlBarrak
- Development and Deployment of an e-Health System in UNRWA Healthcare Centers (HCs): The Experience and Evidence 1212
Ghada Ballout, Najeeb Al-Shorbaji, Amin Shishtawi, Fathia Abuzabaida, Yousef Shahin, Ali Khader, Wafa'a Zeidan, Yassir Turki, Ahmad Al-Natour, Rawan Saadeh, Majed Hababeh and Akihiro Seit
- Cultural Problems Associated with the Implementation of eHealth 1213
H.-F. Mandirola Brioux, S. Benitez, C. Otero, D. Luna, J.H.B. Masud, A. Marcelo, M. Househ, C. Hullin, C. Villalba, S. Indarte, S. Guillen, P. Otero, F. Campos, A. Baum and F. Gonzalez Bernaldo de Quirós
- Management of In-Field Patient Tracking and Triage by Using Near-Field Communication in Mass Casualty Incidents 1214
Po-Liang Cheng, Yung-Cheng Su, Chung-hung Hou and Po-Lun Chang
- Development of Day Care Management Systems for Elderly 1215
Ming-Chu Chiang, Yu-Ting Yeh and Chien-Tsai Liu
- Use of Person-Generated Health Data in Kinect-Based Stroke Rehabilitation Systems: A Systematic Review 1216
Gerardo Luis Dimaguila, Kathleen Gray and Mark Merolli
- A Serious Game for Anterior Cruciate Ligament Rehabilitation: Software Development Aspects and Game Engagement Assessment 1217
Marcos Cordeiro d'Ornellas, Alison Santos Machado, Jefferson Potiguara de Moraes and Ana Lúcia Cervi Prado
- Comparing the Human Papillomavirus Vaccination Opinions Trends from Different Twitter User Groups with a Machine Learning Based System and Semiparametric Nonlinear Regression 1218
Jingcheng Du, Jing Huang, Rui Duan, Yong Chen and Cui Tao
- Social Media as a Communication Support for Persons with Mild Acquired Cognitive Impairment: A Social Network Analysis Study 1219
Aboozar Eghdam, Ulrika Hamidi, Aniko Bartfai and Sabine Koch
- Rapid Development and Distribution of Mobile Media-Rich Clinical Practice Guidelines Nationwide in Colombia 1220
José F. Flórez-Arango, M. Sriram Iyengar, Indira T. Caicedo and German Escobar
- The Diffusion of mHealth Applications 1221
Parisis Gallos and John Mantas
- Warfarin Guide: Co-Design of a Mobile Computer-Assisted Anticoagulant Therapy 1222
Martin Gerdes, Santiago Martinez, Berglind Smaradottir, Rune Fensli and Jarle Jortveit
- Patient-Clinician Co-Design Co-Participation in Design of an App for Rheumatoid Arthritis Management via Telehealth Yields an App with High Usability and Acceptance 1223
Rebecca Grainger, Hermaleigh Townsley, Tobias Langlotz and William Taylor
- Competencies for Effective Use of Online Depression Information Among College Students 1224
Yulong Gu, Zornitsa Kalibatseva and Xu Song

Flexible Detection of Fall Events Using Bidirectional EMG Sensor <i>Hao Han, Xiaojun Ma and Keizo Oyama</i>	1225
Design of Rehabilitation Treatment Coach Robot <i>Seung-Ho Han, Han-Gyu Kim, Chan-Yong Park and Ho-Jin Choi</i>	1226
A Patient Centric Nursing Mobile App Redesign <i>Hernán Navas, Mariano Dozo, Ezequiel Dourado, Fernando Canteros and Ingrid Rung</i>	1227
Connecting PHRs and EHRs for a Sustainable National Health System <i>Evelyn Hovenga and Heather Grain</i>	1228
Intelligent Digital Environment for Wellbeing and Health <i>Inga Hunter, Caroline Lockhart, Phoebe Elers, Dick Whiddett and Hans Guesgen</i>	1229
The Development of mWellcare, an mHealth System for the Integrated Management of Hypertension and Diabetes in Primary Care <i>Devraj Jindal, Priti Gupta, Dilip Jha, Vamadevan S. Ajay, Pramod Jacob, Kriti Mehrotra, Pablo Perel, Jonathan Nyong, Ambuj Roy, Nikhil Tandon, Dorairaj Prabhakaran and Vikram Patel</i>	1230
How Are Doctors Using Mobile Electronic Medical Records? An In-Depth Analysis of the Usage Pattern <i>Juntae Kim, Yura Lee, Sanghee Lim and Jae-Ho Lee</i>	1231
A Content Analysis of Mobile Apps for Chronic Kidney Disease Patient Care: Searching in English and Chinese <i>Ying-Li Lee, Yan-Yan Cui and Polun Chang</i>	1232
Developing Mobile Health Management System for Patients with Musculoskeletal Tumor <i>Nai-Wen Liang, Hung-Yi Liu, Chun-En Chien, Qi-Guang Chen, Po-Lun Chang, Wei-Ming Chen and Bo-Gui Wu</i>	1233
Setting the Agenda for Personal Health Records in Low- and Middle-Income Countries <i>Thomas Mwogi and Martin C. Were</i>	1234
Empathy Bot: Conversational Service for Psychiatric Counseling with Chat Assistant <i>Kyo-Joong Oh, DongKun Lee, ByungSoo Ko, Jonghwan Hyeon and Ho-Jin Choi</i>	1235
Factors Influencing Progress of Health Information Exchange Organizations in the United States <i>Lauren M. Overhage, Jennifer Covich-Bordenick, Xiouchun Li and J. Marc Overhage</i>	1236
3D Documentation of Chronic Wounds Using Low-Cost Mobile Devices <i>Ekaterina Sirazitdinova and Thomas M. Deserno</i>	1237
A Document-Based EHR System That Controls the Disclosure of Clinical Documents Using an Access Control List File Based on the HL7 CDA Header <i>Toshihiro Takeda, Kanayo Ueda, Akito Nakagawa, Shirou Manabe, Katsuki Okada, Naoki Mihara and Yasushi Matsumura</i>	1238
School-Based Telemedicine: Perceptions About a Telemedicine Model of Care <i>May Lin Tye, Michelle Honey and Karen Day</i>	1239
CLARUS as a Cloud Security Framework: e-Health Use Case <i>David Vidal, Santiago Iriso and Rafael Mulero</i>	1240
An Interdisciplinary Approach Between Medical Informatics and Social Sciences to Transdisciplinary Requirements Engineering for an Integrated Care Setting <i>Jan Vielhauer and Britta Böckmann</i>	1241
Use Contexts and Usage Patterns of Interactive Case Simulation Tools by HIV Healthcare Providers in a Statewide Online Clinical Education Program <i>Dongwen Wang</i>	1242
A Homemade Easy Information System: The Example of Patient Satisfaction Survey <i>Hang Yin, Pingping Yan, Yanyan Cui, Polun Chang and Jun Lei</i>	1243
Personal Health Self-Management in a Data Perspective <i>Yiqin Yu, Xiang Li, Gang Hu, Haifeng Liu, Jing Mei, Yuan Ni, Guoyu Tang, Guotong Xie and Weiming Xu</i>	1244
Nurses' Interest, Readiness and Absorptive Capacity to Information Technology: A Survey in China <i>Boning Zhang, Ping Liu and Qian Xiao</i>	1245

II. Health Data Science

A Deep Learning Approach to Neuroanatomical Characterisation of Alzheimer's Disease <i>Abhinith Kumar Ambastha, Tze-Yun Leong and The Alzheimer's Disease Neuroimaging Initiative</i>	1249
Three Dimensional Virtual Planning Through Cone Beam Computed Tomography for Surgical Guidance Production <i>Kaline Thumé Antunes, Marcos Cordeiro d'Ornellas, Daniani Deliberalli Noedel, Gustavo Nogara Dotto and Renésio Armino Ghres</i>	1250
"Big Clinical Data" Analysis of Intravenous to Oral Conversions of Non-Antimicrobials <i>Patrick E. Beeler, David W. Bates and Jürg Blaser</i>	1251
MetaMap Lite in Excel: Biomedical Named-Entity Recognition for Non-Technical Users <i>Ravi Teja Bhupatiraju, Kin Wah Fung and Olivier Bodenreider</i>	1252
An Exploratory Analysis of Questions Submitted to a Brazilian Telemedicine System <i>Jacqueline Brixey and Juliana J. Brixey</i>	1253
Integrating x4T-EDC into an Image-Portal to Establish an Ophthalmic Reading Center <i>Philipp Bruland, Ulrich Kathöfer, Maximilian Treder, Nicole Eter and Martin Dugas</i>	1254
Technical Challenges and Opportunities when Implementing Pharmacogenomics Decision Support Integrated in the Electronic Health Record <i>Pedro J. Caraballo, Joseph A. Sutton, Ann M. Moyer, David Blair, Lois C. Hines, Padma S. Rao, Mark F. Adams, Sahana Murthy, Tina Garza, Mary E. Karow, Harmanjit Singh, Jyothsna Giri, Donald B. Gabrielson, Jennifer St. Sauver, Suzette J. Bielinski and Mark A. Parkulo</i>	1255
Validated Simulation: The Preliminary Experience of Anesthesiologist Board Examination in Taiwan <i>Hung-Wei Cheng, Pin-Tarnng Chen, Ying-Chih Lo and Po-Lun Chang</i>	1256
How to Use TCM Informatics to Study Traditional Chinese Medicine in Big Data Age <i>Cheng Shi, Qing-yue Gong and Jinhai Zhou</i>	1257
Identifying Use Cases for Electronic Nursing Record Systems Using Clinical Workflow Observations and a Delphi Survey <i>Insook Cho and Joohyun Chung</i>	1258
Delivering Digital Drugs: An Exploratory Study of the Digitalisation of Supply and Use of Medicines <i>Tony Cornford, Valentina Lichtner, Jane Dickson, Ralph Hibberd, Ela Klecun, Will Venters and Bryony Dean Franklin</i>	1259
Concept Embedding for Relevance Detection of Search Queries Regarding CHOP <i>Yihan Deng, Lukas Faulstich and Kerstin Denecke</i>	1260
Classifying Clinical Notes with Pain Assessment <i>Samah J. Fodeh, Dezon Finch, Lina Bouayad, Stephen Luther, Robert D. Kerns and Cynthia Brandt</i>	1261
Modeling an IT Support for Handling Serious Adverse Events in Clinical Trials <i>Simon Fonck and Thomas Deserno</i>	1262
A Study on Data-Driven Novel Cancer Staging Methods <i>Yuan Gao, Yu Tian, Shengqiang Chi, Yao Lu, Xinhang Li, Tianshu Zhou and Jing-song Li</i>	1263
Methods for Faster and Efficient Data Entry in Electronic Medical Records <i>Shashi B. Gogia and Anand N. Malaviya</i>	1264
Comparative Analysis of Geriatric and Adult Drug Clinical Trials on ClinicalTrials.gov <i>Zhe He and Aisha Langford</i>	1265
Generation of openEHR Test Datasets for Benchmarking <i>Samar El Helou, Tuukka Karvonen, Goshiro Yamamoto, Naoto Kume, Shinji Kobayashi, Eiji Kondo, Shusuke Hiragi, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda</i>	1266
Mining Associations Between Genes and Clinical Conditions of Breast Cancer by Using Gene Expression Data <i>Yuefang Hou, Hao Zhang, Han Zhang and Lei Cui</i>	1267
Comparing Different Adverse Effects Among Multiple Drugs Using FAERS Data <i>Jing Huang, Xinyuan Zhang, Jingcheng Du, Rui Duan, Liu Yang, Jason H. Moore, Yong Chen and Cui Tao</i>	1268

Usability Evaluation of NLP-PIER: A Clinical Document Search Engine for Researchers <i>Gretchen Hultman, Reed McEwan, Serguei Pakhomov, Elizabeth Lindemann, Steven Skube and Genevieve B. Melton</i>	1269
Explore Care Pathways of Colorectal Cancer Patients with Social Network Analysis <i>Tianyao Huo, Thomas J. George Jr., Yi Guo, Zhe He, Mattia Prosperi, François Modave and Jiang Bian</i>	1270
Constructing an Open-Access Bio-Signal Repository from Intensive Care Units <i>Doyeop Kim, Sukhoon Lee, Tae Young Kim, Sanghyung Jin, JaeYeon Park, JeongGil Ko, Rae Woong Park and Dukyong Yoon</i>	1271
Depression, Impulse Control Disorder, and Life Style According to Smartphone Addiction <i>E.Y. Kim, S.W. Joo, S.J. Han, M.J. Kim and S.Y. Choi</i>	1272
A Performance Comparison on the Machine Learning Classifiers in Predictive Pathology Staging of Prostate Cancer <i>Jae Kwon Kim, In Hye Yook, Mun Joo Choi, Jong Sik Lee, Yong Hyun Park, Ji Youl Lee and In Young Choi</i>	1273
Avoiding Overfitting in Deep Neural Networks for Clinical Opinions Generation from General Blood Test Results <i>Youjin Kim, Han-Gyu Kim, Zhun Li and Ho-Jin Choi</i>	1274
Li2b2-Façade: Simulation of i2b2 Data Warehouse Server and Client for Interaction with Other Systems <i>Raphael W. Majeed, Tingyan Xu, Mark R. Stöhr and Rainer Röhrig</i>	1275
Asynchronous Query Distribution Between Multiple i2b2 Research Data Warehouses: Li2b2-SHRINE <i>Raphael W. Majeed, Mark R. Stöhr, Volker S. Thiemann, Rainer Röhrig and Andreas Günther</i>	1276
Deep Diabetologist: Learning to Prescribe Hypoglycemic Medications with Recurrent Neural Networks <i>Jing Mei, Shiwan Zhao, Feng Jin, Lingxiao Zhang, Haifeng Liu, Xiang Li, Guotong Xie, Xuejun Li and Meilin Xu</i>	1277
System Development Aiming Efficient Recording of Pedigree Information and Database Construction for Genetic Counseling Clinic <i>Takuya Nagasawa, Mitsuhiro Azuma, Michihiko Aki, Masaki Yamada, Mineko Ushiana, Noriko Tanabe, Takeshi Nakajima, Kokichi Sugano, Hiromi Sakamoto and Teruhiko Yoshida</i>	1278
Estimating Sample Size for a Feasibility Study of Computer-Assisted Input Support to EDC <i>Naoki Nishimoto, Naomi Tamura, Jun Kunikata, Sumiko Akahori, Tomoaki Sogo, Yuma Tani and Hideto Yokoi</i>	1279
Extracting Predictive Indicator for Prognosis of Cerebral Infarction Using Machine Learning Techniques <i>Yasunobu Nohara, Koutarou Matsumoto and Naoki Nakashima</i>	1280
Identifying Patients' Smoking Status from Electronic Dental Records Data <i>Jay Patel, Zasim Siddiqui, Anand Krishnan and Thankam Thyvalikakath</i>	1281
Discovery of Psychoactive Substance Addiction Patterns Based on Information Mining Engineering <i>Ramón García-Martínez, Sebastián Martins, Santiago Bianco and Hernán Navas</i>	1282
Bringing Knowledge to Users in One Click: Infobuttons in the Problem List of an EHR <i>Jose F. Rodriguez, Emilio S. Mole, Luciana Rubin, Sonia E. Benitez, Carlos M. Otero, Daniel R. Luna and Fernan Quiros</i>	1283
Effects of e-Learning and m-Learning on Nursing Care in a Continuing Education Context: An Overview of Mixed Method Systematic Reviews (Protocol) <i>Geneviève Rouleau, Marie-Pierre Gagnon, José Côté, Emilie Hudson, Julie Payne-Gagnon, Julien Bouix-Picasso and Carl-Ardy Duboi</i>	1284
Using Human Phenotype Ontology for Phenotypic Analysis of Clinical Notes <i>Feichen Shen, Liwei Wang and Hongfang Liu</i>	1285
Using a Simulation Modelling Approach to Manage Outpatient Department Waiting Time at the National Hospital of Sri Lanka <i>S.N. Silva, R.H. Hewapathirana and W.M.A.U. Jayatilleke</i>	1286
Data Mining Applied to Analysis of Contraceptive Methods Among College Students <i>Priscyla Waleska Simões, Samuel Cesconetto, Eduardo Daminelli Dalló, Maria Marlene de Souza Pires, Eros Comunello, Felipe Borges Tomaz, Eduardo Picolo Xavier, Pedro Antonio da Rosa Brunel Alves, Luciane Bisognin Ceretta and Sandra Aparecida Manenti</i>	1287

Identifying Chemical-Disease Relationship in Biomedical Text Using a Multiple Kernel Learning-Boosting Method <i>Yueping Sun, Yu Zhang and Jiao Li</i>	1288
Automated Report Generation for Research Data Repositories: From i2b2 to PDF <i>Volker S. Thiemann, Tingyan Xu, Rainer Röhrig and Raphael W. Majeed</i>	1289
Extraction and Evaluation of Medication Data from Electronic Dental Records <i>Yue Wang, Zasim Siddiqui, Anand Krishnan, Jay Patel and Thankam Thyvalikakath</i>	1290
Development of Asian Non-Small Cell Lung Cancer Survival Prediction Model Using an Innovative Method of Bayesian Network <i>Bincan Yin, Shichao Xin and Yuhong Zhao</i>	1291
Comparison of Different Algorithms for Sentiment Analysis: Psychological Stress Notes <i>Sunmoo Yoon, Faith Parsons, Kevin Sundquist, Jacob Julian, Joseph E. Schwartz, Matthew M. Burg, Karina W. Davidson and Keith M. Diaz</i>	1292
The Study of Hot Spots on Hepatitis B Dissertation Based on Co-Word Analysis in China <i>W. Zhang, Y.B. Wang, X.Z. Zhang and H.M. Duan</i>	1293
III. Human, Organizational, and Social Aspects	
Nursing Informatics Pioneers Embrace Social Media <i>Juliana J. Brixey and Susan K. Newbold</i>	1297
Virtual Reality in Nursing: Nasogastric Tube Placement Training Simulator <i>Kup-Sze Choi</i>	1298
Digital Health Information Workforce Planning Through a National Census <i>Kathleen Gray and Kerryyn Butler-Henderson</i>	1299
Development of a Data Collection Tool for MbHIS-QUAL: Evaluation of the Quality of Morbidity Data in Routine Health Information Systems (RHISs) in Hospitals <i>Lyn Hammer, Edward Nicola and Debbie Bradshaw</i>	1300
Digital Divide and Health Disparities in China: Data from a National Longitudinal Survey of CHARLS <i>Y. Alicia Hong, Zi Zhou and Ya Fang</i>	1301
Patient Handovers – Cognitively Demanding: Does the Handover EHR Meet This Challenge? <i>Ursula Hübner and Mareike Przysucha</i>	1302
Evaluation of the Anonymity and Utility of De-Identified Clinical Data Based on Japanese Anonymization Criteria <i>Eizen Kimura, Satoshi Hasegawa, Koji Chida, Shoko Gamo, Satoshi Irino, Haku Ishida and Yukio Kurihara</i>	1303
Developing an Evidence-Based Web-Education Program Designed to Help Nurses Learn to Care for Acutely Ill Intoxicated Patients in Emergency Departments <i>Marita Koivunen, Eija Noppari, Tiina Hakala and Anna-Kaisa Kautto</i>	1304
Privacy Policy Implementation on the Nation-Wide EHR in Japan for Hospitals and Patients <i>Naoto Kume, Shinji Kobayashi, Kenji Araki and Hiroyuki Yoshihara</i>	1305
Patient's Adherence Level Determination System <i>Taranik Maksim and Georgy Kopanitsa</i>	1306
Report and Analysis of Web Portal oldagesolutions.org for Indian Old Age Persons <i>Sushil Kumar Meher and A.B. Dey</i>	1307
Authenticating Unknown Doctors for Access to EHRs Based on Societal Trust <i>Kensuke Morris, Naoto Kume, Goshiro Yamamoto, Shinji Kobayashi, Kazuya Okamoto, Hiroshi Tamura and Tomohiro Kuroda</i>	1308
DOctimer: A Timing and Event Recording Tool for Direct Observational Research <i>Matthew E. Nolan, Brian W. Pickering and Vitaly Herasevich</i>	1309
Health Digital Divide in the Era of Precision Healthcare: Taking Yunnan Province as an Example <i>Yan Peng, Li Yan, Liuwang Kuanghua and Guozheng Tang</i>	1310
Measuring User Satisfaction with Clinical Information Systems: What Really Matters? <i>Jonathan L. Schaffer, Peter Haddad and Nilmini Wickramasinghe</i>	1311

Prerequisites for International Exchanges of Health Information for Record Research: Comparison of Australian, Austrian, Finnish, Swiss, and US Policies	1312
<i>Hanna Suominen, Henning Müller, Lucila Ohno-Machado, Sanna Salanterä, Günter Schreier and Leif Hanlen</i>	
Key Factors for the Successful Adoption of IS/IT in Healthcare: A Fit-Viability Perspective	1313
<i>Nilmini Wickramasinghe, Peter Haddad and Imran Muhammad</i>	
Mobile Learning in Nursing Undergraduates in China: Current Status, Attitudes and Barriers	1314
<i>Qian Xiao, Qiannan Zhang, Lanlan Wang, Yanling Wang, Liu Sun and Ying Wu</i>	
Secure Electronical Communications and Data Transfers in a Clinical Environment	1315
<i>T. Xu, V.S. Thiemann, R. Röhrig, D. Brammen and R.W. Majeed</i>	
IV. Knowledge Management	
Big Data in the ICU: Experience in the Hospital Italiano de Buenos Aires	1319
<i>José Astudillo, Javier Zelechower, Francisco Traversaro, Francisco Redelico, Daniel Luna, Fernan Quiros, Marcelo Risk and Eduardo San Román</i>	
Training of Students for Critical Evaluation of Mobile Health Applications	1320
<i>Jadranka Bozikov and Mira Hercigonja Szekeres</i>	
Developing the First Generally-Available openEHR Archetypes and Templates for Physiotherapy: An Example of Building Clinical Models and Modelling Capacity via Student-Led Academic-Industrial Collaboration	1321
<i>Jamila Chihab, Hildegard Franke, Ian McNicoll and Matthew W. Darlison</i>	
Exploiting Temporal Constraints of Clinical Guidelines by Applying OpenEHR Archetypes	1322
<i>Lilian Mie Mukai Cintho, Diego Garcia, Bruno Henrique da Silva Santos, Lucia Sacchi, Silvana Quaglini and Claudia Maria Cabral Moro</i>	
Towards the Consideration of Diagnostic Delay in Model-Based Clinical Decision Support	1323
<i>Jan Gaebel, Mario A. Cypko and Steffen Oeltze-Jafrá</i>	
Considerations of Human Factors in the Design and Implementation of Clinical Decision Support Systems for Tumor Boards	1324
<i>Carmelo Gaudioso and Peter Elkin</i>	
Usability Study of RSNA Radiology Reporting Template Library	1325
<i>Yi Hong, Ying Zhu, Selen Bozkurt, Jin Zhang and Charles E. Kahn Jr.</i>	
Evaluation of SNOMED CT as a Reference Terminology for Standardized Data Queries in the Arden Syntax	1326
<i>Robert A. Jenders</i>	
Building Interoperable FHIR-Based Vocabulary Mapping Services: A Case Study of OHDSI Vocabularies and Mappings	1327
<i>Guoqian Jiang, Richard Kiefer, Eric Prud'hommeaux and Harold R. Solbrig</i>	
Mapping HL7 CDA R2 Formatted Mass Screening Data to OpenEHR Archetypes	1328
<i>Shinji Kobayashi, Naoto Kume and Hiroyuki Yoshihara</i>	
Implementation of an Archetype Data Set to Reuse Electronic Health Record Data in Clinical Decision Support Systems	1329
<i>Georgy Kopanitsa</i>	
Correcting Ontology Errors Simplifies Visual Complexity	1330
<i>Hao Liu, Ling Zheng, Yehoshua Perl, Yan Chen and Gai Elhanan</i>	
Using Semantic Technologies to Extract Highlights from Care Notes	1331
<i>Vanessa Lopez, Grace McCarthy, Joao Bettencourt-Silva, Marco Sbodio, Natasha Mulligan, Fabrizio Cucci, Stephane Deparis, Ciaran Hennessy, Nagesh Yadav, Karie Kelly, Russell Olsen, Conor Cullen and Spyros Kotoulas</i>	
Challenges in Archetypes Terminology Binding Using SNOMED-CT Compositional Grammar: The Norwegian Patient Summary Case	1332
<i>Luis Marco-Ruiz and Rune Pedersen</i>	
Terminology Status APIs – Mapping Obsolete Codes to Current RxNorm, SNOMED CT, and LOINC Concepts	1333
<i>Lee Peters, Thang Nguyen and Olivier Bodenreider</i>	

From a Categorical Structure to a Concept Model: The International Classification of Health Interventions (ICHI)	1334
<i>Jean-Marie Rodrigues, Sukil Kim, Syed Aljunid, Jae Jin Lee, Huib Ten Napel and Béatrice Trombert</i>	
POLE.VAULT: A Semantic Framework for Health Policy Evaluation and Logical Testing	1335
<i>Arash Shaban-Nejad, Anya Okhmatovskaia, Eun Kyong Shin, Robert L. Davis and David L. Buckeridge</i>	
Clinical Workflow Modeling in Obstetrics: Hepatitis B in Pregnancy	1336
<i>Fadi Shamoon, Harald Leitich, Jeroen S. de Bruin, Andrea Rappelsberger and Klaus-Peter Adlassnig</i>	
CoMetaR: A Collaborative Metadata Repository for Biomedical Research Networks	1337
<i>Mark R. Stöhr, Gudrun Helm, Raphael W. Majeed and Andreas Günther</i>	
Using an Ontology-Based Approach to Handle Author Affiliations in a Large Biomedical Citation Database	1338
<i>Haixia Sun, Junlian Li, Yingjie Wu, Lei Wang and Kin Wah Fung</i>	
Self-Reported eHealth Literacy Among Undergraduate Nursing Students in Selected Districts of Sri Lanka	1339
<i>Shaluni Tissera and Nishan Silva</i>	
The Analysis of Medical Adverse Events Related to Electronic Health Records in Nursing Services	1340
<i>Mika Tomita, Natsumi Kishi and Mariko Iwasawa</i>	
Using SNOMED Distance to Measure Semantic Similarity of Clinical Trials	1341
<i>Duo (Helen) Wei and Gang Fu</i>	
Terminology Adoption in Family Medicine Service – Start of Journey	1342
<i>Austen Wong, Maggie Lau, Vicky Fung and Dorothy To</i>	
Application of Information Technology in the Outpatient Service Optimization	1343
<i>Xiaoying Zhang</i>	
An Open Metadata Schema for Clinical Pathway (openCP) in China	1344
<i>Wei Xu, Yanxin Zhu and Xia Wang</i>	
Integration of Japanese Medical Device Adverse Event Terminologies	1345
<i>Ayako Yagahara, Takumi Tanikawa, Katsuhiko Ogasawara and Hideto Yokoi</i>	
Infrastructure for Big Data in the Intensive Care Unit	1346
<i>Javier Zelechower, José Astudillo, Francisco Traversaro, Francisco Redelico, Daniel Luna, Fernan Quiros, Eduardo San Roman and Marcelo Risk</i>	
Information Model Construction of EBM Resources	1347
<i>Yanxin Zhu, Wei Xu and Xia Wang</i>	
V. Quality and Safety, and Patient Outcomes	
A Hospital Nursing Adverse Events Reporting System Project: An Approach Based on the Systems Development Life Cycle	1351
<i>Yingjuan Cao and Marion Ball</i>	
Data Quality of the Chinese National AIDS Information System: A Critical Review	1352
<i>Hong Chen, Ping Yu, David Hailey and Tingru Cui</i>	
Evaluation of Clinical Nursing Information System in Taiwan Regional Hospital	1353
<i>Wan-Jie Chou, Pei-Yu Tsai, Shu-Ying Lin and I-Ching Hou</i>	
Personalizing Longitudinal Care Coordination for Patients with Chronic Kidney Disease	1354
<i>Theresa A. Cullen, Suranga N. Kasthurirathne, Jenna M. Norton and Andrew S. Narva</i>	
A Guide for Constructing Bayesian Network Graphs of Cancer Treatment Decisions	1355
<i>Mario A. Cypko, Matthaeus Stoehr, Steffen Oeltze-Jafra, Andreas Dietz and Heinz U. Lemke</i>	
Modeling a System for Generating Structured Reports	1356
<i>Lorena Dutra da Costa and Marcos Cordeiro d'Ornellas</i>	
Event Notification in Support of Population Health: The Promise and Challenges from a Randomized Controlled Trial	1357
<i>Brian E. Dixon and Kenneth S. Boockvar</i>	

Automatic Loading of Problems Using a Comorbidities Subset: One Step to Organize and Maintain the Patient's Problem List	1358
<i>Santiago Márquez Fosser, Alejandro Gaiera, Carlos Otero, Sonia Benitez, Daniel Luna and Fernán Quiroz</i>	
Using Health Information Technology to Enhance Care Outcome Accountability Through Bundled Payments	1359
<i>Rima Artonian Gibbings and Nilmini Wickramasinghe</i>	
Can Video Conferencing Facilitate Better Discharge Processes and a Superior Patient Experience?	1360
<i>Jeremy Grummet, Nilmini Wickramasinghe, Peter Haddad and Louise O'Connor</i>	
A New Essential Functions Installed DWH in Hospital Information System: Process Mining Techniques and Natural Language Processing	1361
<i>Masayuki Honda and Takehiro Matsumoto</i>	
A Study of Evaluation and Automation Instruction of Management by Law of Community Health Service in China	1362
<i>Hongpu Hu, Yue Zhang, Yao He, Fan Yu and Minjiang Guo</i>	
Using a Mobile Device Application to Support Emergency Clinicians in Diagnosing Pulmonary Embolism	1363
<i>Chun Yen Huang, Po Liang Cheng and Po Lun Chang</i>	
The Impact of PACS on Radiograph Interpretations in an Orthopedic Outpatient Clinic	1364
<i>Petter Hurlen, Arne Borthne and Pål Gulbrandsen</i>	
Comparing Inpatient Falls Guidelines to Develop an ICNP®-Based Nursing Catalogue for ENRs	1365
<i>Miran Jung, Insook Cho, Jisun Chae, Jihye Kim and Yeunhee Kim</i>	
Willingness to Pay for Elderly Tele-Care System Using Digital Terrestrial Broadcasting	1366
<i>Ogasawara Katsuhiko, Kaga Satoshi, Suzuki Teppei, Tani Yuji, Higashikawa Fumihiro, Enomoto Hisashi, Moriyama Hiroyuki and Kise Nobuyuki</i>	
Development of a Quick SOFA-Based Sepsis Clinical Decision Support System in a Tertiary Hospital Emergency Department	1367
<i>Youn-Jung Kim, Jae-Ho Lee, Sang-Wook Lee, Ming-Young Jung and Won Young Kim</i>	
Development of Safety and Usability Guideline for Hospital Information System	1368
<i>Yura Lee, Min-Young Jung, Gee Won Shin, Sangwoo Bahn, Taezoon Park and Jae-Ho Lee</i>	
The Role of Technology in Medication Safety Incidents: Interpretative Analysis of Patient Safety Incidents Data	1369
<i>Valentina Lichtner, David Gerrett, Ann Slee, Noreen Gul and Tony Cornford</i>	
Using Modified-ISS Model to Evaluate Medication Administration Safety During Bar Code Medication Administration Implementation in Taiwan Regional Teaching Hospital	1370
<i>Pei-Luen Ma, Yan-Wun Jheng, Bi-Wei Jheng and I-Ching Hou</i>	
Decision Support System for Medical Care Quality Assessment Based on Health Records Analysis in Russia	1371
<i>Maksim Taranik and Georgy Kopanitsa</i>	
Method to Identify Diagnostic Rules for Pancreatic Cancer Using Laboratory Data Based on Bayesian Estimation	1372
<i>Shirou Manabe, Yoshie Shimai, Taizo Murata, Ayumi Fujii, Kanayo Ueda, Akito Nakagawa, Toshihiro Takeda, Naoki Mihara and Yasushi Matsamura</i>	
The Evaluation of the Health Information Exchange with the Number of Usage and the Introduction of Outpatient to the Hospitals at Nagasaki Japan	1373
<i>Takehiro Matsumoto and Masayuki Honda</i>	
Computerized Cognitive Rehabilitation: Comparing Different Human-Computer Interactions	1374
<i>Silvana Quaglini, Anna Alloni, Barbara Cattani, Silvia Panzarasa and Caterina Pistarini</i>	
In Vitro Comparative Study Between Conventional and Computer-Assisted Surgery Methods for Planning and Resection of Bone Sarcomas	1375
<i>L.E. Ritacco, L.A. Aponte-Tinao, D.L. Muscolo, M.A. Ayerza, I. Albergo, G.L. Farfalli, M. Risk, F. Gonzalez Bernaldo de Quiros and F.E. Milano</i>	
Effects of Computerized Guideline-Oriented Clinical Decision Support System on Glycemic Control in Diabetic Patients: A Systematic Review and Meta-Analysis	1376
<i>Ryota Sakurai and Kazuhiko Ohe</i>	
Use of an Off-the-Shelf Corporate Information Tool to Track a High-Level-Disinfection Process	1377
<i>Carolina Salazar, Jaime de los Hoyos, Francisco Izurieta, Daniel Jara and Daniel Capurro</i>	

Using Fast Healthcare Interoperability Resources (FHIR) for the Integration of Risk Minimization Systems in Hospitals <i>Kutaiba Saleh, Stephan Stucke, Alexandr Uciteli, Sebastian Faulbrück-Röhr, Juliane Neumann, Kais Tahar, Danny Ammon, Tobias Schmidt, Thomas Neumuth, Andreas Besting, Frank Portheine, Heinrich Herre, André Kaeding and Martin Specht</i>	1378
Development of Observation Support System Using Integrated Nursing Practice Data <i>Ryoma Seto, Toshitaka Inoue, Nobuyuki Koeda, Sayuri Yamashita, Kotaro Komino and Takahiro Yasoshima</i>	1379
Accuracy of Decision Support Systems for Breast Cancer – Initial Results <i>Priscyla Waleska Simões, Ronaldo Borges Vicente, Patricia Duarte Simões Pires, Maria Marlene de Souza Pires, Eros Comunello, Felipe Borges Tomaz, Luciane Bisognin Ceretta and Deborah Ribeiro de Carvalho</i>	1380
Accessibility of Mobile Devices for Visually Impaired Users: An Evaluation of the Screen-Reader VoiceOver <i>Berglind Smaradottir, Jarle Håland and Santiago Martinez</i>	1381
Advancing Towards a CDA-Based Trauma Registry Data Submission <i>Martin Staemmler, Alexander Nischwitz, Markus Blätzing and Johannes Sturm</i>	1382
Patient Accessibility to Hospitals in Winter Road Conditions: GIS-Based Analysis Using Car Navigation Probe Data <i>Takumi Tanikawa, Hisateru Ohba, Ayako Yagahara and Katsuhiko Ogasawara</i>	1383
Implementation of an Outsourced Transnational Service of Clinical Decision Support System <i>Vania Teixeira, Luciana Rubin, Romina Rebrij, Alvaro Tamborindeguy, Rodrigo Martinez, Juan Carlos Bacigalupo and Daniel Luna</i>	1384
Methods for Detecting Malfunctions in Clinical Decision Support Systems <i>Adam Wright, Trang T. Hickman, Dustin McEvoy, Skye Aaron, Angela Ai, Joan S. Ash, Jan Marie Andersen, Rachel Ramoni, Milos Hauskrecht, Peter Embi, Richard Schreiber, Dean F. Sittig and David W. Bates</i>	1385
Data Mining on Numeric Error in Computerized Physician Order Entry System Prescriptions <i>Xue Wu and Changxu Wu</i>	1386
Development of Multiple Wireless Network Environments for Hospitals <i>Yoshinori Yamashita and Tatoku Ogaito</i>	1387
Core Curriculum Comparisons in Master Degree Programs for Patient Safety <i>QiYan Zhang and Hiroshi Takeda</i>	1388
Subject Index	1389
Author Index	1395

PAPERS

This page intentionally left blank

I. Connected and Digital Health

This page intentionally left blank

Multimodal e-Health Services for Smoking Cessation and Public Health: The SmokeFreeBrain Project Approach

Panagiotis D. Bamidis, Evangelos Paraskevopoulos,
Evdokimos Konstantinidis, Dimitris Spachos, Antonis Billis

Lab of Medical Physics, Medical School, Aristotle University of Thessaloniki, Greece

Abstract

Smoking is the largest avoidable cause of preventable morbidity worldwide. It causes most of the cases of lung cancer and chronic obstructive pulmonary disease (COPD) and contributes to the development of other lung diseases. SmokeFreeBrain aims to address the effectiveness of a multi-level variety of interventions aiming at smoking cessation in high risk target groups within High Middle Income Countries (HMIC) such as unemployed young adults, COPD and asthma patients, and within the general population in Low-Middle Income Countries (LMIC). The project addresses existing approaches aimed to prevent lung diseases caused by tobacco while developing new treatments and evaluating: (i) Public Service Announcement (PSA) against smoking, (ii) the use of electronic cigarettes, (iii) neurofeedback protocols against smoking addiction, (iv) a specifically developed intervention protocol based on behavioral therapy, social media/mobile apps and short text messages (sms) and (v) pharmacologic interventions. Emphasis in this paper, however, is placed on the e-health, m-health, open (big) data, mobile game and neuroscientific challenges and developments upon facilitating the aforementioned interventions.

Keywords:

Smoking Cessation, Telemedicine, Neuroscience

Introduction

Smoking is the largest avoidable cause of preventable morbidity and premature mortality worldwide [1]. The prevalence of smoking worldwide is estimated at about one billion smokers, half of whom will die prematurely as a consequence of their addiction, unless they quit [2]. Smoking causes approximately 85% of the cases of lung cancer and chronic obstructive pulmonary disease (COPD) and contributes to the development of many other lung diseases [3]. Therefore, the control of smoking and the active reduction of exposure to tobacco substances in the environment are considered highly important interventions for lung disease prevention [4].

According to World Health Organization (WHO) one third of the world adult population today, 1.1 billion people are smokers and 3.5 million deaths per year are attributed to smoking. It is estimated that in 2020 the number of deaths per year will be increased to 10 million. Smokers are 13 times more likely to die from COPD than nonsmokers.

Tobacco consumption is highly influenced by socioeconomic factors, affecting mostly low- and middle-income countries as well as vulnerable populations within high income countries. Additionally, smoking causes health inequality between

gender and age groups [4], while it significantly elevates the preventable morbidity and premature mortality rates worldwide.

In the past, our team has described an approach towards the development of an integrated system supporting the smoking cessation network initiatives in Greek public hospitals. The system combined the availability of an open source, web based EHR subsystems, with a Web 2.0 facilitated e-learning component for supporting continuing medical education and promoting public awareness [5].

With these in mind, the SmokeFreeBrain project (www.smokefreebrain.eu), funded by the H2020 programme of the European Commission, aims to address the effectiveness of a multi-level variety of intervention strategies for smoking cessation in high risk target groups within High Middle Income Countries (HMIC) such as unemployed young adults, chronic obstructive pulmonary disease (COPD) and asthma patients, and within the general population in Low Middle Income Countries (LMIC). As the main effect of the recent economic crisis in various European countries, is a significant increase of youth unemployment. The present project focuses on how this socioeconomic development affects the vulnerability of the young population. The project addresses existing approaches aimed to prevent and control lung diseases caused by tobacco consumption, while proposing the development of new treatments and analyzing their contextual adaptability to the local and global health care system as it is affected by the recent socioeconomic changes [6].

In this paper, the various SmokeFreeBrain interventions are first outlined. Emphasis is then placed on the facilitation of these interventions with various e-health, m-health, open (big) data, mobile game and neuroscientific developments where technical aspects such as the design and system architectures of the database registry, as well as, the mhealth systems, together with the informatics implications for pilot preparations are discussed.

SmokeFreeBrain Objectives and Concept

The objectives of the project are outlined below:

- Examine the effects of the use of electronic cigarettes during the initial phase of smoking cessation along with the possible formation of carcinogenic nitrosocompounds via the exposure to nicotine through electronic cigarette vaping.
- Examine the global DNA methylation status under two different situations, tobacco smoking and inhaling e-cigarette vapor.

- Development and evaluation of a novel neurofeedback protocol for smoking cessation.
- Develop a smoking cessation intervention based on adherence to physical activity with ICT support (App Gamification, Facebook and SMS).
- Generate and validate a set of software tools that can be used to inform EU policymakers and local governments as to how to produce optimal Public Service Announcements (PSA) regarding smoking.
- Develop a best practice guide for promoting smoking cessation and how these can be applied in large scale
- Evaluate and report on the effectiveness of the proposed interventions in terms of cost, socioeconomic and health demographics terms and report to policy making bodies.

SmokeFreeBrain follows an interdisciplinary approach in various relevant fields in order to generate new knowledge. State of the art techniques in toxicology, pulmonary medicine, neuroscience and behavior are utilized to evaluate the effectiveness of: (i) Public Service Announcement (PSA) against smoking, (ii) the use of electronic cigarettes with and without nicotine as a harm reduction approach and/or cessation aid, (iii) a specifically developed neurofeedback intervention protocol against smoking addiction, (iv) a specifically developed intervention protocol based on behavioral therapy, social media/mobile apps and short text messages (sms) and (v) pharmacologic interventions.

In attempting to set a summarised outline of the project in a nutshell, one may say that it will use samples of the populations showing high vulnerability to smoking in order to study the effects of five distinct interventions for smoking cessation to measure their cost-effectiveness and allow conclusions regarding the outcomes scalability in terms of health economics. Figure 1 illustrates this concept.

SmokeFreeBrain Interventions

e-Cigarette Intervention

Electronic cigarettes (EC) are a product that can be used for consumption of nicotine-containing vapour via a mouth piece, or any component of that product, including a cartridge, a tank and the device without cartridge or tank. Electronic cigarettes can be disposable or refillable by means of a refill container and a tank, or rechargeable with single use cartridges. ECs or electronic nicotine delivery systems (ENDS) are also defined as devices whose function is to vaporize and deliver to the lungs of the user a chemical mixture typically composed of nicotine, propylene glycol and other chemicals, although some products claim to contain no nicotine. These products are not currently regulated or monitored. As such, contents may vary and may not be known to the consumer. The safety of ECs has not yet been scientifically demonstrated. Additionally the products vary widely in the amount of nicotine and other chemicals they deliver and there is no way for consumers to find out what is actually delivered by the product they have purchased. Consumers often believe that the use of ECs is safer than smoking tobacco. The chemicals used in ECs have not been fully disclosed, and there are no adequate data on their emissions [7].

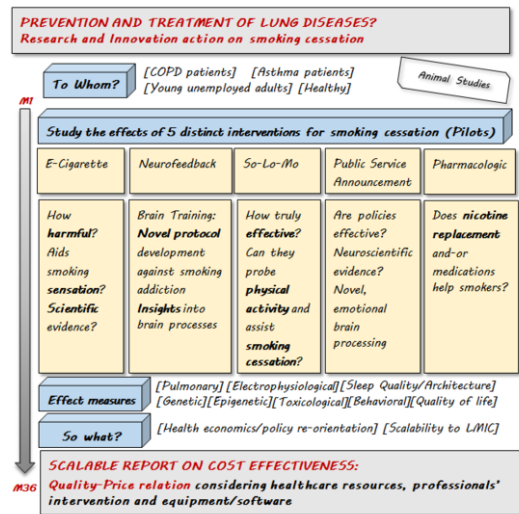


Figure 1 – A conceptual outline of the SmokeFreeBrain project Neurofeedback Intervention

As a non-pharmacological, non-invasive and painless brainwave technique, Neurofeedback (NF) trains individuals to take control of their mind through operant conditioning [8]. As nicotine consumption acts on the brain through specific receptors causing changes in neurological function, the goal of NF interventions is for an individual to learn to control his or her brain function without the need of relevant psychoactive substances. Therefore, NF can eliminate addictive behaviors and lead to better results for smoking cessation than other mainstream approaches, as recent studies indicate. Although a universally accepted neurofeedback protocol for smoking cessation therapy is not yet established, several studies employing real time fMRI (rtfMRI) revealed that reductions in activity of specific brain regions and craving have been observed [9].

Social Media: So-Lo-Mo (Social-Local-Mobile) Intervention

There is some evidence both on the usefulness of exercise for smoking cessation and the motivating effect of Social Networks with successful people trying to quit smoking using the same process. Therefore, an intervention based on adherence to physical activity with ICT support (App Gamification, Facebook and SMS) may provide a valuable aid for smoking cessation [10]. In addition, craving is one key component that has been shown to vary over time during a smoking-cessation attempt and to be highly related to treatment efficacy and cessation success [10]. The objective of the So-Lo-Mo intervention would be twofold. First, the So-Lo-Mo intervention will focus on distracting the user by delivering several tasks through a mobile app. By taking into consideration the understanding of craving during smoking cessation attempts and based on the user's profile, So-Lo-Mo asks the user to undertake some short-term or long-term tasks. The short-term tasks rely on specific physical exercise activities (utilizing the smartphone's accelerometer), gamification, socialization, goal assignment and achievements list, etcetera. The long-term tasks are represented by physical exercise.

Public Service Announcements (PSAs)

Governments across the world are required to disseminate information concerning risks to public health and to promote messages that encourage healthier life style options to improve public health and reduce the huge burden placed on state spending from state subsidized health care in countries with social security systems such as Germany, the UK and France. In this context, PSAs are non-commercial advertisements intended to achieve attitudinal and behavioural changes in the public. PSAs are at the core of many public health campaigns against smoking, and other possible public health problems.

When effective, PSAs are of substantial benefit to public welfare. However, the lack of reliable, quantitative and objective means of evaluating advertising effectiveness is one of the key impediments to better PSA outcomes. In addition, poorly designed PSAs can often have effects that are contrary to their desired goals [11]. Over the past few years, considerable developments in methodologies to record brain activity (e.g. functional Magnetic Resonance Imaging (fMRI), Electro-Encephalography (EEG)) and psychophysics have allowed novel paradigm designs and analyses providing key insight into both the explicit (conscious) and implicit (subconscious) cerebral responses to PSAs. This intervention uses modern neuroscientific tools with the latest thinking and methodological approaches available in the fields of marketing and social science and the interaction of experts from academia, private companies and governmental organizations to derive a new device that can measure and predict the efficacy of PSAs related to smoking abuse.

Pharmacologic interventions

This intervention will investigate the effects of current pharmacological treatments on sleep quality and physiology of COPD and asthma patients as well as in young unemployed adults focusing on two well-known drugs: Varenicline and bupropion. Varenicline is a chemical substance that it used as medication for smoking cessation and affects the nervous system by making nicotine less effective in two different ways: by acting like nicotine as a partial agonist to reduce craving for cigarettes or by replacing nicotine (agonist) and decreases the pleasurable effect of cigarettes. However, there are studies that report an adverse drug reaction of "abnormal sleep related events" associated with varenicline. To allow for such inferences, SmokeFreeBrain will conduct a sleep study, which will be performed using overnight polysomnographic (PSG) recordings consisting of electroencephalography (EEG), electrooculography (EOG), electrocardiography (ECG) and electromyography (EMG) sensors.

The insights gained from the pharmacological studies will generate new knowledge regarding the toxicity of these interventions according to their effectiveness. These effects will then be compared with the toxicity of e-cigarettes via that intervention, allowing inferences regarding the safety of the relevant, corresponding use.

ICT positioning of the project

SmokeFreeBrain will develop and deploy both cutting edge and stable, well-understood, technologies (from previous EU projects and commercial partners) within the new field of smoking cessation support, including neurofeedback intervention, gamification platform, mobile app for delivering challenges and interaction information, serious games promoting physical exercise, neurometric approaches, and decision support for challenging and motivational tasks and

smoking related profile databases to maximise value from research and to optimise commercial opportunities. Technical partners have assessed technology readiness levels of existing components to be used in SmokeFreeBrain. Figure 2 illustrates the current technology readiness levels (TRLs) in line with the corresponding estimated readiness level at the end of the project (TRLs vary from 1 to 9).

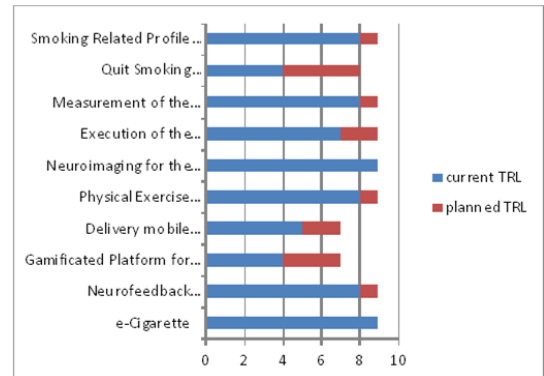


Figure 2 – TRLs of SmokeFreeBrain technologies at project start and estimates envisioned at the end of the 3 year run.

Data Management Plan

The data management plan will guide all activities regarding the anonymization, exchange and release of data gathered in the pilot sites as required for participation in the Open Research Data Pilot in H2020. Datasets that are produced within the SmokeFreeBrain project span from smokers' demographic and clinical data to outcome measures of several clinical interventions and their cost-effectiveness results. These data will allow for the benchmarking of the different approaches when dealing with smoking cessation, providing a common basis for further policy decision making. Since SmokeFreeBrain pilots involve human participants, data collected will contain sensitive personal information. Focus is also given to possible ethical issues and access restrictions regarding personal data so that regulations on sensitive information are not violated.

The data management portal will be based on the popular open source software CKAN and it will be accessible through a portal (endpoint) at the following address: ckan.smokefreebrain.org

CKAN is a powerful data management system that makes data accessible by providing tools to streamline publishing, sharing, finding and using data. CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available. The software supports strong integration with third-party CMS's, such as Drupal and WordPress.

Among the portal's features will be:

- Complete catalog system with an easy to use web interface and a powerful API
- Integration with the profile database system
- Data visualization and analytics
- Fine access controls
- Raw data and metadata storage

- Search by keyword or filter by tags. See dataset information at a glance.
- Rich application programming interface (API), and over 60 extensions including link checking, comments, and analytics and many more.

The SmokeFreeBrain Profile database

The profile database will be a clinical trial database. It will be a general data hub for the needs of the project and will hold data from all interventions. Partners can use the profile database with 2 ways:

- As a database for their intervention
- As a hub for storing data from interventions for processing purposes

In the first case, partners will be able to insert their data into the database. In the second case, they have to send the data via web services, which will follow the ISO EN 13606 standard. A general overview of the system architecture can be shown in Figure 3. The profile database will produce reports and statistics based on the stored data. The reports will be used to evaluate the results of the interventions of the project.

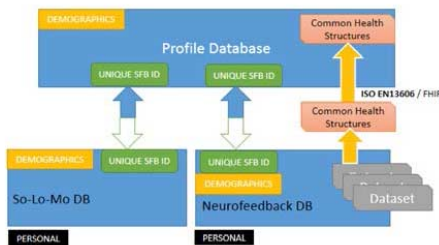


Figure 3 – A general overview of the SmokeFreeBrain system.

The SmokeFreeBrain API

This API is used to Post, Read and Update data in the Profile Database using the CEN/ISO EN13606, a European norm from the European Committee for Standardization (CEN) which is also approved as an international ISO standard [12].

The SmokeFreeBrain SoLoMo App

The main features of the SoLo-Mo app are:

- Personal profile
- Achievements: Cigarettes not smoked, money savings, hours of live regained, time smoke free
- Total fitness time (Google Fit integration): Daily average, monthly average, days above the average
- Self-help contents
- Motivational tools: Readings, relaxation tool and mini-games
- Relapse tracker
- Encouraging messages: Both from the system, and from/to other users

The SmokeFreeBrain SoLoMo App mini-games

An interesting feature of the SoLo-Mo app is the concept of mini-games. The relevant contribution of mini-games in

smoking cessation has already been recognized in recent studies [13]:

- Playing Tetris decreases craving strength for drugs
- Puzzle-solving games reduce craving for nicotine
- Breath-control App emulating the physiological responses smokers get from smoking
- Crushing virtual cigarettes contributes to smoking cessation

We focused on 3 mini games:

- Crushing cigarettes [14]
- Breathing control by blowing up balloons
- Physical Exercise Coach/Instructor (based on the webFitForAll exer-gaming platform protocol) [15, 16]

For example, in the *Blowing up Balloons* mini-game, emphasis is placed on breathing control. The user is asked to blow close to the microphone when the balloon is green. If the user blows when the balloon is red, the balloon may pop. If the user does not blow when the balloon is green, the balloon deflates quickly. The number of totally blown balloons in 1 min reflects the total score, while the pop balloons are counted too. This is based on the Jamalian et al. conceptual study:

- Phone app that emulates the physiological responses smokers get from smoking
- Innovative breath-control element, which mimics the physiological and perceived effects of nicotine as well as the relaxant effects of smoking.
- The goal is to give players a cigarette-free, non-invasive way to satisfy their nicotine cravings

Figure 4 provides a rough screenshot of the game interfaces.



Figure 4 – A sense of the Blowing up Balloons SmokeFreeBrain and SoLo-Mo app minigame.

Envisioning the SmokeFreeBrain stakeholder ecosystem

The contribution of SmokeFreeBrain towards the creation of a social ecological model relies on understanding the multiple stakeholders and their needs. SmokeFreeBrain will develop and support a collaborative health-ecosystem by bringing together the relevant key stakeholders who will testify on their changing interactions and endorse the implemented interventions to ensure maximum health benefits. At the policy level, alliance with national healthcare policy makers will capitalize efforts and scale up interventions in a global context. The organisational stakeholders are hospitals, health service providers and national authorities who monitor existing approaches to prevention or develop treatments. The community stakeholders include the scientific community (e.g. researchers, academics), commercial actors

(pharmaceutical industry, insurers) and cross-sectoral experts (e.g. researchers on environmental effects). The interpersonal level includes the So-Lo-Mo community, existing smoking cessation networks, family and friends. The patient/smoker is central to the entire ecosystem. SmokeFreeBrain will inform all levels of the framework by providing access to real and valuable information on the establishment of effective interventions for the prevention and treatment of lung diseases. Figure 5 illustrates the above described ecosystem.

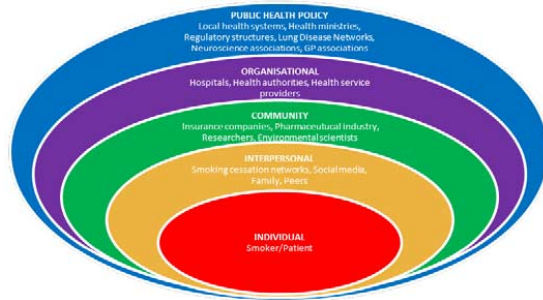


Figure 5– SmokeFreeBrain stakeholder ecosystem.

Conclusion

The SmokeFreeBrain project kicked off in November 2016 and is expected to have a three year duration. At the point of writing this paper, numerous multi-centric trials are underway. These are expected to shape the way forward by providing innovative systems, interventions and new, original, knowledge derived as statistical analysis and economic evaluation will soon kick-off. This paper has attempted to provide some overall, but technically oriented glimpses on the project concept. Given space limitations and the early stages of the trials, more details and results were thought improper for the purposes of this paper. Nevertheless, the technical frontiers which have been outlined in this paper can not be understated.

Acknowledgements

The SmokeFreeBrain project (www.smokefreebrain.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 681120.

References

- [1] WHO, Priority Medicine for Europe and the World 2013 Update (http://www.who.int/medicines/areas/priority_medicines/MasterDocJun_e28_FINAL_Web.pdf)
- [2] World Health Organization. Tobacco- Fact Sheet No339, July 2013. Available from: <http://www.who.int/mediacentre/factsheets/fs339/en/> (accessed 21 September 2013).
- [3] Lung cancer- Cancer Research UK. Available from: <http://www.cancerresearchuk.org/cancer-help/type/lung-cancer/>
- [4] The World Health Report 2002—Reducing Risks, Promoting Healthy Life (<http://www.who.int/whr/2002/en/>)
- [5] Konstantinidis ST, Konstantinidis E, Nikolaidou MM, Boutou AK, Havouzis N, Argyropoulou P, Bamidis PD. The use of open source and Web2.0 in developing an integrated EHR and e-learning system for the Greek Smoking Cessation Network. *Stud Health Technol Inform.* 2009;150:354-8.
- [6] T. McAfee and Others Helping Smokers Quit — Opportunities Created by the Affordable Care Act *N Engl J Med* 372:5, 2015
- [7] Position Statement on Electronic Cigarettes [ECs] / Electronic Nicotine Delivery Systems (ENDS)
- [8] Hammond, D. C. (2011). What is neurofeedback: An update. *Journal of Neurotherapy*, 15(4), 305-336.
- [9] Hanlon, C. A., Hartwell, K. J., Canterberry, M., Li, X., Owens, M., LeMatty, T., & George, M. S. (2013). Reduction of cue-induced craving through realtime neurofeedback in nicotine users: the role of region of interest selection and multiple visits. *Psychiatry Research: Neuroimaging*, 213(1), 79-81.
- [10] King, D., Greaves, F., Exeter, C., & Darzi, A. (2013). 'Gamification': Influencing health behaviours with games. *Journal of the Royal Society of Medicine*, 106(3), 76-78.
- [11] Biener, L., & Siegel, M. (2000). Tobacco marketing and adolescent smoking: more support for a causal inference. *American journal of public health*, 90(3), 407.
- [12] SmokeFreeBrainproject deliverable D2.5 App for Patient, Medical Station and So-Lo-Mo Administration Console, available through www.smokefreebrain.eu
- [13] Skorka-Brown J, Andrade J, Whalley B, May J, Playing Tetris decreases drug and other cravings in real world settings. *Addictive Behaviors* 51 (2015) 165–170.
- [14] Girard B, Turcotte V, Bouchard S, Girard B. Crushing virtual cigarettes reduces tobacco addiction and treatment discontinuation. *Cyberpsychol Behav.* 2009 Oct;12(5):477-83. doi: 10.1089/cpb.2009.0118.
- [15] Konstantinidis EI, Billis AS, Mouzakidis CA, Zilidou VI, Antoniou PE, Bamidis PD. *IEEE J Biomed Health Inform.* 2016 Jan;20(1):189-200. doi: 10.1109/JBHI.2014.2378814.
- [16] Bamparopoulos, G., Konstantinidis, E., Bratsas, C., & Bamidis, P. D. (2016). Towards exergaming commons: composing the exergame ontology for publishing open game data. *Journal of biomedical semantics*, 7(1), 1.

Address for correspondence

Panagiotis D. Bamidis,
Lab of Medical Physics, Medical School,
Aristotle University of Thessaloniki,
PO Box 376, 54124,
Thessaloniki, Greece,
tel: +30-2310-999310,
email: pdbamidis@gmail.com

Using Smartphones for Research Outside Clinical Settings: How Operating Systems, App Developers, and Users Determine Geolocation Data Quality in mHealth Studies

AL Beukenhorst^a, DM Schultz^b, J McBeth^a,
R Lakshminarayanan^c, JC Sergeant^{a,d}, and WG Dixon^a

^a Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

^b Centre for Atmospheric Science, School of Earth and Environmental Sciences, University of Manchester, Manchester, UK

^c uMotif, London, UK

^d NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, Manchester, UK

Abstract

Smartphones that collect user geolocation provide opportunities for mobile Health (mHealth). Although granularity of geolocation data may be high, data completeness depends on the device's operating system, application developer decisions, and user actions. We investigate completeness of geolocation data collected via smartphones of 5601 people that self-reported daily chronic pain symptoms on 349,293 days. On 17% of these days, hourly geolocation data is reported, but days with 0 (16%), 1 (14%) and 2 (13%) geolocations are common. Android phones collect geolocation more often than iPhones (median 17 versus 2 times a day). Factors on operating system level and individual user level influence completeness of geolocation data collected with smartphones. mHealth researchers should be aware of these factors when designing their studies. The mHealth research community should devise standards for reporting geolocation data quality, analysing systematic differences in data quality between participant groups, and methods for data imputation.

Keywords:

Mobile Health (mHealth); Smartphone; Longitudinal Studies

Introduction

The high penetration of smartphones in our society enables large-scale collection of self-reported data in real-time for research outside clinical settings. Moreover, these temporally-rich data sets can be complemented with passively collected data by smartphone sensors, such as physical activity, sleeping patterns or geolocation data [1]. Geolocation data is especially useful as it can be linked to openly available data sets with high spatial and temporal resolution (e.g. climate, weather, air pollution).

Usage of smartphones to collect geolocation data passively has many advantages. It reduces the burden of participation as participants can use their own device and do not need to manually enter location data. It is less prone to recall bias compared to keeping location diaries, increasing the internal validity of studies [2]. Through the Global Positioning System (GPS), location can be sampled both frequently and accurately, rendering assumptions that participants always stay at their home or work postcode unnecessary.

Although granularity of geolocation data may be higher using smartphones, completeness and accuracy of passively collected sensor data may be less when using smartphones rather than clinically-approved devices and more variable between participants who use different smartphones. The accuracy and completeness of geolocation data are affected by a variety of factors relating to operating system [3], application (app) developer decisions [4; 5] and smartphone settings defined by the user [4; 5], which are partially outside researchers' control. Thus, if geolocation completeness patterns systematically differ between participant groups, the validity of study results may be compromised.

The purpose of this study is to assess completeness of geolocation data collected during the study Cloudy with a Chance of Pain. This is a national UK smartphone study investigating the association between the weather and chronic pain. Study participants rate ten aspects of their daily symptoms in the study's smartphone app, which collects geolocation data to retrieve hourly weather data from the nearest weather station via the Met Office DataPoint service.

We investigated data completeness for the study population and compared the number of geolocation observations between Android and iPhone users, as well as their demographics.

Sampling geolocation data: location services on Android and iOS smartphones

The availability of geolocation data depends on the smartphone's operating system, various settings defined by the app developer, and phone settings chosen by the user [4; 5]. The "Cloudy with a Chance of Pain" app was developed for the two operating systems with largest market share in the UK [6]: Android (53% market share), developed by Google and used on many different brands of phones, and iOS (44% market share), developed by Apple for iPhones only. We will first discuss settings that affect availability and accuracy of geolocation data independent of operating system, then discuss the operating system-specific settings under various developers' decisions and finally the user-defined settings. Choices made during development of the Cloudy with a Chance of Pain app are listed in the Methods section below.

Both operating systems can employ the GPS, network signals (a combination of WiFi and cell-tower signals, and, in some cases, signals of so-called Bluetooth beacons), or both to determine the smartphone user's geolocation. These strategies

vary in accuracy, availability and battery power consumption. GPS provides the highest accuracy, but only works outside of buildings and uses high battery power. Deriving a user's location based on network signals usually has a lower accuracy, but consumes less battery power and is available both inside and outside. Battery consumption also depends on frequency of location requests, the requested accuracy and the frequency of uploading location data to the application's servers. App developer guides for both Android and iOS applications encourage app developers to choose settings that preserve battery power.

In both operating systems, apps can be 'active' (in use on the smartphone's screen), 'in the background' (not in use, but not terminated), or 'not running' (not launched or terminated by system/user).

Location services in Android

For Android, Google's mobile operating system, app developers specify the logic for receiving location updates: what strategy to use (GPS, network or both), frequency, accuracy and the thresholds for requesting updates (time interval or change in distance). The app collects multiple location updates from the GPS and/or Network Location Provider for a specified amount of time, and then chooses the best estimate. The logic for best estimate is determined by the app developer, depending on the purpose of the app (e.g. most recent, most accurate). Usage of the GPS provider and size of window for collecting location updates are associated with highest power consumption.

In Android, location updates are available when the app is running, when the app is in the background, but not when the app is terminated. If a user 'restricts background data', location services (and other functionalities) are suppressed when a user is not connected to WiFi. Retrieval of location updates can automatically be resumed when the device is restarted [4].

Location services in iOS

For iOS, Apple's mobile operating system for iPhones, three strategies for determining user location exist: standard location service (GPS), the significant-change location service (network signals, but only supplied after a developer-specified 'significant change' in user's location) and region monitoring (detects user's entry and exit into specific regions). App developers specify the strategy, frequency of updates and thresholds for requesting updates.

Compared to Android, iOS highly regulates geolocation requests when the app is in the background. Background processes are always time-restricted, making continuous data collection almost impossible. After an app is terminated, GPS geolocation is not available, but the significant-change location service can restart the app 'in the background' every fifteen minutes, reporting location changes as long as the operating system allows. If 'Background App Refresh' is disabled by the user, none of the location services are available when the app is not in use. Location updates can also be programmed by the developer to pause under certain conditions (e.g. when the user is unlikely to be moving) [5].

User actions affecting location services

The smartphone user can also influence availability of geolocation data. On both operating systems, users can (a) disable location services for all apps in the general settings, (b) deny location services for the specific app, (c) switch off the device, (d) switch the device to a mode that does not allow location services (i.e. airplane mode, or, for Android phones,

battery-saving mode), or (e) switch off WiFi/mobile network (only affects geolocation through network signals).

Methods

Participants

UK residents who experienced pain for at least the preceding 3 months were invited to participate, through charity and patient organisations, and through study publicity on television, radio and in newspapers from January 20th 2016. Further requirements for participation were: Age 17 or over, resident in the UK and owning an Android (Android 3 or later) or iOS smartphone (iOS 8 or later). To enrol in the study, participants downloaded the uMotif Cloudy with a Chance of Pain app, consented to participate and completed a baseline questionnaire.

Ethical approval was obtained in December 2015 from the University of Manchester Research Ethics Committee 4 (ref: ethics/15522).

Cloudy with a Chance of Pain app

Every day, participants received a reminder to rate ten aspects of their symptoms in the app on a five-point ordinal scale. They could answer (any of) the ten aspects multiple times a day, for example in case of changing pain. In parallel, the smartphone's GPS passively recorded geolocation up to hourly, thus ideally recording 24 geolocation observations per day.

A participant-day was defined as a day on which the application collected any data (i.e. symptom rating submitted by the participant, passively recorded geolocation, or both). We distinguished between participant-days with and without symptom data. Participant-days eligible for the main analysis were those on which a participant rated at least one of the ten symptom aspects; days on which a participant did not record any self-reported data, but the app did collect geolocation, were excluded.

The Android and iPhone app were developed to be as similar as possible in their geolocation retrieval strategy. The app uses both GPS (outdoors) and network signals (inside buildings). Every hour on the hour, it retrieves the last observed geolocation in the previous 60 minutes and its accuracy. If only network signals are available, the location update is only triggered *in case of significant change* (i.e. if a user is inside a building and does not move for an hour, the event is not fired). Geolocations and their timings are saved by the app until participants have internet connection, and then sent to the servers.

Analysis

Participants eligible for this analysis were those recruited between 20 January and 16 November 2016 and who had provided symptom data for at least 7 separate days.

We analysed data completeness for all participant-days that met the inclusion criteria. We compared data completeness between Android and iOS participant-days, and the demographic characteristics between the Android and iOS users. The Mann-Whitney *U* test was used to test whether the number of geolocations per participant-day differed between operating systems and whether the number of symptoms per participant-day differed between operating systems.

Results

Subjects

Figure 1 shows the flowchart of study participants. A total of 12,441 people downloaded the smartphone application. We excluded participants who did not provide any symptom data at all ($N = 1222$), that provided symptom data for less than 7 days ($N = 5383$) or for whom we could infer that they manually disabled location services for our app, never allowing geolocation requests ($N = 205$). Of the 5601 remaining participants, 47% used an Android device, 17% used an iOS device and for 37% the device type could not be determined because the users enrolled in the study before reporting of operating-system type was added to the app's functionalities.

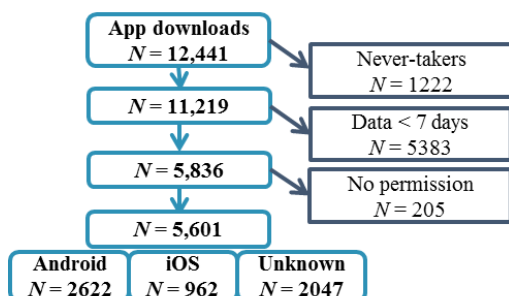


Figure 1 – Flowchart of number of participants. We excluded participants who never submitted symptom data, symptom data on less than 7 days or that did not give allow geolocation requests.

Participant-days

Figure 2 shows the participant-days. The 5601 included participants provided data on 492,080 days. We excluded

Participantdays - total:

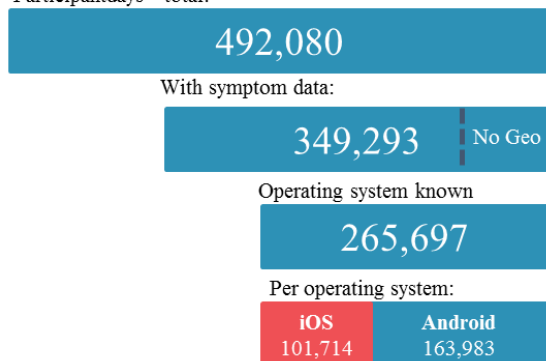


Figure 2 – Participant-days of the 5601 included participants.

142,787 participant-days on which geolocation was collected, but participants did not provide symptom data. Of the remaining 349,293 days, 28% were provided by the iOS users, 46% were provided by the Android users and 26% were provided by users for which operating system was unknown. On average, participants submitted symptom data on 62 days. The median number of days was 33 (IQR: 14–76).

Overall completeness of geolocation data

Figure 3 shows the distribution of completeness of geolocation data for days on which users submitted symptom data. Of all

the participant-days, 16% had no geolocation data. Of the remainder, the most common number of geolocations was 24 (17%), followed by 1 (14%), then 2 (13%). Figure 4 shows the distribution of geolocation data completeness by known device type for all days of participants, as percentage of participant-days per device type. Most of the 101,714 participant-days collected by iOS devices collected either two geolocations (27%) or one geolocation (26%). For the 163,983 participant-days collected by Android

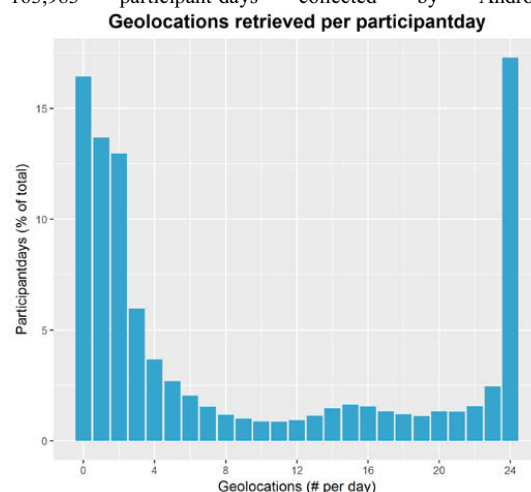


Figure 3 – Distribution of number of geolocations retrieved on 349,293 participant-days with symptom data.

Geolocations retrieved per participantday by operating system

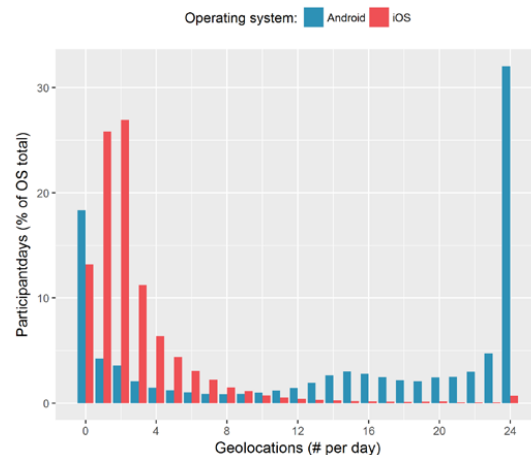


Figure 4 – Distribution of number of geolocations, retrieved on 265,697 participant-days for Android users (blue bars; $N = 163,983$ participant-days) and 972 iOS users (red bars; $N = 101,714$ participant-days) that did not disable location services over the complete study period. Participant-days with an unknown operating system are excluded.

devices, the most common were 24 geolocations (32%), followed by zero geolocations (18%). Table 1 shows that the median number of GPS observations was 17 for Android participant-days and 2.8 for iOS participant-days. The Mann-Whitney U test showed a statistically significant difference in distribution of number of geolocations per participant-day

between both operating systems ($p < 0.01$). However, the difference in number of symptoms submitted by the participants (mean 10.4 versus 10.3) is not practically meaningful. This difference suggests that an average Android user submits one complete symptom rating more than an iOS user after 100 days of symptom data entry. The statistically significant difference in symptom ratings between the two operating systems can be explained simply by the large volume of data.

Table 1 – Summary of data completeness of geolocation and self-reported data completeness per participant-day for Android and iOS users. Number of symptom data entries per participant-day refers to the number of symptom aspects rated by that participant (one complete submission consists of 10 symptom aspects).

		Android <i>N</i> = 163,983	iOS <i>N</i> = 101,714
Geolocation	Mean	14.1	2.8
	Median	17	2
Symptoms	Mean	10.4	10.3
	Median	10	10

Discussion

We investigated data completeness of a large mobile Health (mHealth) study in which participants rated ten aspects of their chronic pain in a smartphone app that simultaneously collected their geolocation. Although the app was developed to retrieve 24 geolocations per day (i.e. hourly), it was only successful on 17.3% of the participant-days. Part of the missing data can be explained by users switching to airplane mode or battery-saving mode during the day. Users may also have manually terminated the app, in which case geolocation updates are only triggered in case of ‘significant change’. Participants are unlikely to make these significant location changes when they are sleeping, and, when they are at work for those with sedentary jobs. Another possible explanation is the timing of the prompt to submit symptoms – 18:24 in the evening – which was expected to fit best into participants’ lifestyles. If participants manually terminate the app and then stay put for the evening, the location strategies not requiring ‘significant change’ would only be launched after 18:24 in the evening, resulting in missing data in the hours before. Especially interesting is the high prevalence of days with no geolocation updates. Following the app-developers’ settings, we would expect at least one geolocation would be retrieved when the participant submitted symptom data, even if the device had not ‘significantly changed location’ before that time. As many people charge their phones overnight, battery life is expected to decrease over the day, increasing the chance that the phone is in battery-saving mode at 18:24.

Distribution of data completeness differed greatly between the two operating systems. Specifically, an average of 14.1 geolocations was collected on participant-days of Android users, compared to 2.8 for iOS participant-days. As there was no meaningful difference in symptom-entry behaviour between Android and iOS participants, this difference is probably due to the tendency of iOS to block data processing when an app is ‘in the background’. Although percentages of iOS participant-days with few (1–9) geolocation retrievals were higher, a higher percentage of Android participant-days had 0 geolocation retrievals. This may be because historically, only Android phones had a battery-saving mode. This functionality was only introduced in iOS version 9, briefly

before the start of this study (16 September 2015), but cannot be installed on the iPhone 4 and older models.

A limitation of this study was that we had to exclude 2047 participants whose operating system was unknown because they joined the study before the functionality of recognising device type was added to the app. Furthermore, we could not determine the cause of missing geolocation data because our application does not collect information on app status or any user-defined settings.

We have not found mHealth studies that are comparable to Cloudy with a Chance of Pain in number of participants, operating system, and geolocation sampling frequency. The most similar study was a feasibility study of using smartphones to investigate air-pollution exposure [7]. This study had a smaller sample size ($N = 54$) and only developed an application for Android phones. They did not use Android’s positioning system, but used external software called Skyhook to collect geolocation more frequently (every 5 minutes). Their study lasted for three months. Due to differences in reporting, it is hard to compare our study with theirs. For example, they excluded 10 participants that did not keep manual location diaries and 4 participants who did not report having their phone powered on 24 hours a day from the analysis. Of the remaining 38 participants, 4 participants had, on average, complete geolocation data (mean number of geolocations per day 288, every 5 minutes) and another 26 participants had an average of 144 or more (i.e. geolocation every 10 minutes). Only one participant had a daily average under 72 data points, probably still providing more granular data than participants in our study. The authors found that geolocation data completeness depended on smartphone manufacturer, mobile network and the time of day data points were collected, as well as whether the user switched the device off or to airplane mode.

Although data completeness of smartphone-collected geolocation may be suboptimal in our study, various strategies may enable mHealth researchers to still perform the intended investigation. Investigating the richer Android geolocation data to infer information on the sparser iOS data might be possible. By analysing individual patterns in geolocation, recurring periods in which no significant change in location takes place may be identified.

The results of as [7] indicate, however, that even when developing an app for Android users only, data completeness may not be optimal. As iOS’s market share is 44% in the UK (large, compared to Brazil’s 4% and China’s 17% [6]), including both Android users and iOS users in medical studies, may reduce selection bias. On the other hand, any systematic differences between iOS users with sparse geolocation data and Android users with richer geolocation data could also compromise the study’s validity.

In conclusion, this study shows that mHealth researchers collecting geolocation via participants’ own smartphones may still encounter missing data, especially in participants that use iPhones rather than Android devices. As future work we plan to investigate robust methods for imputing geolocation data. Researchers should develop research smartphone applications with the operating systems’ specifications in mind and instruct participants to refrain from configuring their phone to settings that do not allow geolocation requests. In addition, the mHealth research community should devise standards for reporting geolocation data quality and analysing systematic differences between participant groups.

Conclusion

We have shown that a smartphone app, when sampling the user's geolocation at regular intervals, does not always collect hourly geolocation. Missing data is more common in iPhone users compared to Android users.

To increase the data quality of geolocation data, mHealth researchers should be conscious that geolocation availability and accuracy depends upon the operating system, app developer and user. Smartphone applications should be designed with these factors in mind.

Our results indicate that geolocation sampling may also be suppressed by the operating system, especially on iOS devices, leading to missing data outside of researchers' control. Methodologies to assess – and possibly correct for – systematic differences in data quality between groups of participants should, therefore, be devised.

Acknowledgements

The study was supported by Arthritis Research UK (grant reference 21225), the Arthritis Research UK Centre for Epidemiology (grant reference 20380) and the Farr Institute @HeRC (MR/K006665/1). This report includes independent research supported by the National Institute for Health Research Biomedical Research Unit Funding Scheme. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

References

- [1] B.H. Dobkin and A. Dorsch, The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors, *Neurorehabil Neural Repair* **25** (2011), 788-798.
- [2] S. Shiffman, A.A. Stone, and M.R. Hufford, Ecological momentary assessment, *Annu Rev Clin Psychol* **4** (2008), 1-32.
- [3] D. Ben-Zeev, S.M. Schueller, M. Begale, J. Duffecy, J.M. Kane, and D.C. Mohr, Strategies for mHealth research: lessons from 3 mobile intervention studies, *Adm Policy Ment Health* **42** (2015), 157-167.
- [4] <https://developer.android.com/guide/topics/location/strategies.html>, Location Strategies, in.
- [5] <https://developer.apple.com/library/content/documentation/UserExperience/Conceptual/LocationAwarenessPG/CoreLocation/CoreLocation.html>, Core Location, in.
- [6] <http://www.kantarworldpanel.com/global/smartphone-os-market-share/>, Kantar Worldpanel: Smartphone OS Sales Market Share Great Britain, in.
- [7] M.L. Glasgow, C.B. Rudra, E.H. Yoo, M. Demirbas, J. Merriman, P. Nayak, C. Crabtree-Ide, A.A. Szpiro, A. Rudra, J. Wactawski-Wende, and L. Mu, Using smartphones to collect time-activity data for long-term personal-level air pollution exposure assessment, *J Expo Sci Environ Epidemiol* **26** (2016), 356-364.

Address for correspondence

Anna Beukenhorst, Room 2.811, Stopford Building, Faculty of Biology, Medicine and Health, Arthritis Research UK Centre for Epidemiology, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, The University of Manchester, Oxford Road, Manchester, M13 9PT

The Future of Mobile Usability, Workflow and Safety Testing

Elizabeth Borycki^a, Yalini Senathirajah^b, Andre W. Kushniruk^a

^a School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada

^b State University of New York, New York, New York, United States of America

Abstract

In this paper, the authors outline a vision for the future of mobile usability, workflow and safety testing. The authors argue for the use of glasses that can audio and video record usability, workflow and safety data. Here, citizens, patients and health professionals would become collectors of study data as they use mobile devices and software to support healthcare in the real world. This has become quite feasible with the introduction of low-cost glasses and software applications that allow for the uploading of data for additional analysis by researchers and evaluators of mobile technologies.

Keywords:

Patient Safety; Health Information Systems; Medical Errors

Introduction

Mobile healthcare is on the rise. Citizens and health professionals are using mobile phones to support their healthcare anytime and anywhere (i.e. in the hospital, home and community) [1]. Yet, the usability, safety and workflow implications of these devices and their associated software applications has not been fully researched [2-4]. In this vision paper we consider the use of glasses that audio and video record in terms of: (1) usability, workflow and safety testing, (2) citizens as collectors of audio and video data for research, (3) the potential of video and audio recording devices, (4) the methodological issues associated with using glasses that audio and video record, and (5) future research directions.

Background Literature Review

Why is Usability, Workflow and Safety Testing of Mobile Devices and their Software Important?

Around the world, citizens (e.g. patients, caregivers and health professionals) are using mobile devices such as tablets, mobile phones and smart watches to support their health and wellness, as well as to engage in activities related to healthcare (e.g. looking up health research online, recording physiologic data, using mobile software applications) [1]; for example, using mobile healthcare software applications (e.g. exercise, diet and mood monitoring applications) to self-manage their wellness and chronic diseases [1]. This is consistent with Fox and Duggan's research findings that identify 38% of American adults use exercise, fitness, pedometer or heart rate monitoring health applications; 31% use diet, food or calorie counter software applications; and 12% use mobile weight monitoring software applications [5].

In addition to this, citizens are repurposing existing mobile technologies (e.g. text messaging, mobile calling, photo applications), and using social media software applications that are not traditionally thought of as healthcare applications (e.g. Twitter®, Facebook®, YouTube®) to obtain and communicate health and healthcare information and to manage their health (e.g. take a picture of a rash) and during health crises (i.e. to communicate information about a family member who is critically ill) [1,5].

Internationally, there has arisen a recognition that mobile phones and their software applications hold significant promise for improving the health of individuals, patient health outcomes and the health of populations [1]. The potential power of mobile technologies when applied to healthcare is significant when considering statistics that reveal close to a 100% adoption of these technologies in the developed and developing world [5,6]. Attention to the mobile health revolution is needed. Researchers need to understand how mobile technologies and their software are used (i.e. through usability testing) and integrated into everyday and worklife (i.e. healthcare related activities and workflows). There is also a need to assess the safety of these technologies with the introduction of technology that causes medical errors. These conditions need to be satisfied to ensure citizen safety and improve patient and population health outcomes [1-4, 7].

To illustrate, in a recent publication by Li et al. the researchers found that usability testing led to the refinement of the software user interface design and improved clinician workflows. This in turn led to higher rates of adoption of the technology among clinicians [8]. In a study on mobile prescribing systems safety, researchers were able to identify user interface features and functions that led to prescribing errors (i.e. technology-induced errors). In identifying these user interface features and functions and refining the software, potential medical errors were reduced and the costs associated with treating patients who are harmed or injured were also reduced or eliminated [2, 8].

These issues are not limited to software designed for healthcare use. Mobile hardware and software repurposed by citizens and health professionals for health activities can also introduce errors; for example, the autocorrect function on mobile phones may lead to the wrong name of a medication being recorded (if the auto correct changes the name of the medication) while being written into a note taking app on a mobile phone [2,9]. It is for these reasons that usability, workflow and safety of mobile devices and software need to be tested.

What is the Current State of Mobile Device and Software Usability, Workflow and Safety Testing?

Less attention has been paid by researchers and industry to developing new methods for mobile device and software testing for usability, workflow and safety testing [10]. There has also been less attention focused on understanding how mobile technologies' usability, workflow and safety influence citizen and health professional users in the research literature [11]. There are a number of reasons for this. Perhaps one of the biggest barriers is the mobile nature of devices and the limited development of methods that can be used to test mobile devices for usability, workflow and safety in an effective manner [2,9]. There is a need to develop research methods that allow for recording of audio and video data for analysis of the impact of these technologies in the areas of usability, workflow and safety. Once mobile devices have been refined, we will be better able to understand their impacts upon citizen health, patient outcomes and health professional work processes [9].

In-situ Usability Testing

As mentioned earlier, there has emerged an increasing need to evaluate the effects of mobile device hardware and software from a usability and workflow perspective. Such work is critical to fully understand how the technologies are used to support the health of individuals and populations. "In situ" usability testing, which refers to the collection of usability and workflow data in real-world settings, has become important to understanding how: (a) mobile phone software applications, (b) non-healthcare, mobile phone software applications downloaded by citizens and (c) "true" mobile healthcare software applications are used together by citizens in differing contexts to support health and healthcare [10]. Early work done in this area by Kushniruk and colleagues [10], has often involved conducting usability testing in real-world work settings (i.e. at a computer in a hospital or in a physician's office). Here, electronic health records and hardware devices used in a hospital or physician's office are used during usability and workflow testing to ensure the participant feels they are in a real-world environment [10]. Such testing, although effective, is limited by the type of hardware used (i.e. desktop computer or laptop) and is specific to a particular fixed setting by the hardware (i.e. using a desktop computer in a hospital) [11].

With the introduction of mobile phones and software (including healthcare software applications), there has emerged a need to consider and extend existing approaches to usability, workflow and safety testing to account for the ability of citizens and health professionals to move from one setting to another as they perform activities that influence their health (e.g. dietary choices, the choice to exercise) [9]; for example, "in-situ" testing should take place anywhere a citizen may go and use technology to support their health activities [10]. To illustrate, a citizen might use a mobile device and software to support health-related activities in the home, in the car on the way to work, at work, in a restaurant at lunch, at the gym and/or in the physician's office [5]. This represents a significant challenge for researchers: to collect a full-scale dataset that can be used to improve the overall user experience in engaging in healthcare activities and managing their health, while at the same time collecting information about mobile device hardware and software in these real-world settings [2,10,9,12].

Citizens as Collectors of the Data

Around the world, citizens (e.g. patients, caregivers and health professionals) can now post and view pictures and video online, increasing the potential for consuming and exchanging differing types of health information [13,14]. We have seen rapid growth in the use and miniaturization of cameras and audio recorders, and the trend towards supporting citizens, patients and health professionals in recording their own video and audio data (e.g. mobile phones have cameras, video and audio recorders, smart watches have audio recording capabilities, and glasses can be used to record audio and video) for social, recreational, work and educational purposes. [13-18]. Citizens are posting health information online, and this information may include textual, picture, video and audio data [13, 14]. Some of these postings take the form of information posts, such as links to health information and posts about self help strategies. Some posts include video and audio data uploaded to social media such as YouTube® and Snapchat® [13,15].

The potential to involve citizens and health professionals in the collection of video and audio usability data using these technologies, due to their miniaturized form and integration with other technologies (e.g. glasses, mobile phones, smart watches and social media), is significant [15,16]. Yet, there are few papers that describe the potentials and pitfalls of citizens and health professionals using these ubiquitous tools to collect data in usability, workflow and safety testing (i.e. testing for usability, technology-induced errors and health information technology safety). To date, the focus on such testing has been on mobile devices and software in a traditional laboratory setting [2, 9, 11].

The Challenge of Collecting Mobile Video and Audio Data

Audio and video recordings are essential components of usability testing. Audio data allow one to document verbalizations. If an individual is participating in a usability test, then audio recordings allow for collection of "think aloud" verbalizations or the thoughts of the user as they arise while using a technology, such as a mobile healthcare application [2,11]. Video data in usability testing should include collection of computer screen or mobile device screen recordings [2,9,12]. This is important as it is necessary to be able to view how the user is interacting with the user interface of the technology as they verbalize their thoughts [2,12].

The freedom of movement (i.e. physical and contextual) associated with using a mobile device is significant and poses a number of challenges for conducting research. Mobile device users of healthcare applications seldom stay in a fixed context or position when using the technology. The portability of the device allows the citizen or health professional to move from one location to another and to interact with the device while moving in differing locations throughout the day. This must be accounted for in testing studies. Workflows embedded in the software and workflows associated with using the mobile device and software application need to be captured, and this can be challenging in the real world (i.e. in-situ) [2, 3,10,12].

Existing Approaches Toward Mobile Device and Software Testing

Many methods have been used to collect usability data using mobile approaches. Research methods and factors that influence the quality of mobile usability testing data are

important to consider when designing studies. There are a number of challenges associated with conducting usability tests using mobile phones and healthcare applications [2,12]. One such challenge arises from the recording equipment and the mobile device. The quality of the data collected during mobile usability, workflow and safety tests can be influenced by the type of mobile device and recording equipment/software used. This influences the interpretability of the research results. To illustrate, if the mobile phone is used to record mobile phone screens as the user moves through a software application, hand motions and screen touches will not be recorded. This in turn will make it difficult to determine what the user was touching on the screen at what point in time and if the user interface is confusing [9,12].

In such cases an external audio and video recording device is needed to fully understand what the user is touching on the mobile phone user interface. External devices may require that the mobile phone be fixed to a specific location and position so that all user interactions are captured. This may limit the user's movement, as well as the researcher's ability to record the impact of the device and software upon work tasks and activities [2,12].

Some researchers have documented their experiences in conducting usability and workflow testing involving mobile phone hardware and software. In this work the researchers identified a number of issues that need to be considered when conducting such testing, including: data quality issues, the Hawthorne effect and the inability of participants to engage in typical activities and contexts due to the limitations of the audio and video recording equipment employed in mobile phone and software testing. There is a need to attend to the type of audio and video recording equipment used, as this may affect data quality; for example, recording audio and video data directly to the mobile phone of the user does not provide the researcher with insights into what the user is touching on the mobile device screen or what other artifacts and events in the user's environment are being considered or are influencing mobile phone and software use [2,9,12].

Alternatively, if a user is participating in a test in a usability lab, then data can be captured regarding the user's interactions with the mobile device using an external camera, but the user's interaction with their environment is limited to that of the usability lab, rather than the real world (i.e. a citizens home, work or community environment) [10,12]. The user cannot engage in activities that are typical of those undertaken when using a mobile phone in the real world, in-situ setting (e.g. sitting in a restaurant and recording information about one's meal) [12].

The Future: Using Glasses that Audio and Video Record Data

What is the Potential for Glasses as Video and Audio Recording Devices?

The development of glasses as a video and audio recording tool had its origins in the surveillance [19] and sports industries [20]. In the surveillance industry, glasses that can record video and audio data are used by law enforcement officials to collect data in real time from the perspective of the user for the purpose of conducting surveillance of criminal activities [19]. In the commercial sports industry there emerged a demand for a lightweight, unobtrusive method to collect video and audio data for the purpose of providing the

user's view of sports activities, such as snow boarding, skiing, mountain biking and skateboarding [20]. Users have posted videos for family and friends to share their experiences (with those who were unable to participate) [15,20]. Initially, many of these glasses recorded only video data. Today, they record both audio and video data that can be easily downloaded for review from the glasses [14, 15, 19, 20].

The use of wearable audio and video recording devices has increased in popularity with a wider consumer interest in the technology developing since 2013 [14,15,18]. Approaches have emerged that can be used to effectively collect mobile video and audio recordings for the purpose of assessing the usability, workflow and safety of mobile devices. The idea of collecting audio and video data using glasses has only recently been considered as a viable means of collecting usability, safety and workflow data [21]. In 2013, early work in this area was fostered by Google® with the development of Google® glasses that could be used to present the user with information gleaned from the world wide web, while at the same time providing the capacity to audio and video record data [18]. Wider citizen interest in glasses that record audio and video data increased in 2016 when Snapchat® began marketing glasses that record video and audio data from the user's point of view. Snapchat® glass users, after recording videos or taking pictures using the glasses, can post digital images and video online via an online social media tool [14-16].

Viability of Glasses as Video and Audio Recording Devices

The viability of glasses as an audio and video recording tool needs to be considered in the context of citizens and health professionals as collectors of video and audio data for usability, workflow and safety testing. Glasses can be used to record audio and video data to capture user interactions with mobile device hardware, software and the real world. When considering the technology for mobile usability, workflow and safety testing one must consider that the technology is used independently from the mobile phone with the user wearing the device. Here, glasses are used to not only collect data about devices and software, but also about how these technologies are used in the context of the user's surrounding environment [1].

Earlier, the authors of this paper identified there are several types of glasses that perform this activity on the consumer market. Some have been specifically developed for research purposes or for work use (e.g. Tobii®, Google® glasses). Other glasses have been developed for consumer use (to be used by citizens to record events or sport activities, e.g. Snapchat® glasses). Regardless of the type of glasses used to audio and video record data, there are a number of methodological issues that need to be considered, which we will discuss in the next section of this paper [14-20].

Field of View

One critical issue that needs to be considered when using glasses for audio and video recording data is the field of view. The field of view refers to the extent to which the observable world is seen by the user. Humans have a forward facing horizontal diameter of almost 180-degrees in their field of view [15,18]. In selecting glasses for video and audio recording of interactions with a mobile device and the surrounding environment, it is important to select a device that provides a field of view that is representative of what the user would typically see. The more consistent the field of view is with the users, the more representative the data will be. Some glasses reduce the field of view, thereby limiting the amount

of video data that is collected. Some glass manufacturers have attempted to address this issue by providing glasses with a field of view that is more consistent with the typical user's field of view; for example, Snapchat's® glasses have a 115-degree field of view [15]. As the field of view is not consistent across glass types, pilot testing needs to take place to ensure full video data collection.

User Control of Activation/De-activation

One of the key issues in using glasses in video and audio recording data is the activation of the recording function. Different glasses have different approaches to user control of the recording function. This is important if users are asked to wear the device and record interactions. The ability to activate and de-activate the recording function is important from a methodological perspective. Some glasses do this through voice command, others in response to head motions and still others through a touch pad on the glasses themselves [15-20].

In terms of analysis and use of mobile devices for health purposes, the user could activate and de-activate the recording functions of the glasses, recording only when using the mobile device for health-related activities. The advantage of this approach is that the researcher does not have to review hours of video and audio data, and collected data is focused on the activity of interest. To ensure recording device and application use for health-related activities becomes the focus of the study, glass users may need to be trained to record the activities the researcher is interested in studying.

Privacy, Confidentiality and Ethics

From an ethical perspective, the ability to activate or de-activate the recordings at key points in time is critical from a user's perspective. Here, the user could ask others for permission to record an encounter involving the mobile device, a health app and a health professional or caregiver. Similarly, if the user is unable to obtain the consent of others, the recording could be deactivated by the user. Also, there is the ability to deactivate the recording when the user does not want specific video or audio recorded. To illustrate, Google® glasses can be activated/deactivated through the use of voice commands or head movement at key points in time to enable recording of health related activities [17,18]. Snapchat® glasses offer another approach to the issue of recording. Snapchat® glasses can also be activated/deactivated through a mobile Snapchat® app [15,16]. Snapchat® affords the user some additional features that may be of value from an ethical perspective. Snapchat® glasses indicate when the video and audio camera is recording to those in the field of view (i.e. a light on the glasses indicates when the recording is on). This may be of value to those being video recorded as they can see when the recording function is enabled and disabled [15,16].

Sufficient Length of Recording Time

Length of recording time refers to the amount of time that is dedicated to recording audio and video data. The recording time needs to be of sufficient length to be able to fully record a health-related activity involving a mobile device. There is a range in the length of recording times for differing types of glasses. Short recording times may lead to disruptions in user activity [9]. This may lead to segmented disruptions in user data or place a burden periodically on the user of downloading recordings to a software application, leading to missing data if the user forgets to download data periodically.

Cost

Cost is a consideration when conducting studies involving the use of glasses. The cost of purchasing glasses that can record video and audio data for usability testing can range from \$120 to several thousands of dollars [15, 17, 18, 20]. Glasses used to collect data for usability testing at the upper end of this cost range not only record data, but include analysis software (i.e. Tobii®) [20]. However, involving multiple participants may be costly or cumbersome to coordinate if sharing the glasses between users is difficult. Snapchat® has developed low cost glasses that can be used to live stream and record audio and video data for upload. Snapchat® glasses are priced at \$129/pair, and this includes glasses and a charger. Snapchat® offers prescription glass users the opportunity to purchase a pair of glasses with their own prescription [15,16]. Other glasses used for sports and surveillance purposes that can record audio and video data have also entered the consumer market, allowing for audio and video recording for prices as low as \$50, making real-world studies inexpensive and feasible to conduct with large sample sizes [20].

Aesthetics

A key issue associated with the use of glasses for recording of audio and video is aesthetics. Some users of glasses that audio and video record do not want to wear glasses that differ from other types of glasses that citizens currently wear (e.g. sunglasses, prescription glasses) during the study [15,16,20]. This may decrease participation, as Google® glass users are reluctant to continue using the technology due to privacy concerns for others in their field of view.

Future Research Directions and a Future Vision

Glasses that audio and video record data offer significant advantages over traditional recording techniques used to collect usability, workflow and safety data. Such glasses allow one to record the user's interactions with the mobile device hardware and software (i.e. health app and mobile phone app software). In addition to this, the glasses can record user interactions with their environment, in context and how the mobile technology is used within that context, as recorded in the field of view of the glasses.

Historically, glasses that audio and video record were costly (i.e. \$30,000) [18]. The cost of the technology has been significantly reduced in the past few years (i.e. \$50-100) [20]. The technology is now available at low cost, and there has arisen significant consumer demand for video and audio recording glasses for personal use as a result. Citizens are uploading video to social media (e.g. YouTube®, Snapchat®) to share their experiences and perspectives of the world (i.e. their field of view) [14,15]. Citizens are now capable of collecting usability, workflow and safety data about the mobile hardware and software they use for health activities. There is a significant opportunity to employ consumers in the process of collecting research data. As the cost of the technology has decreased, there are opportunities to increase the sample size of usability, safety and workflow studies and for these studies to take place in real-world contexts.

There remain a number of challenges in conducting such research. Researchers need to understand the audio and video recording glasses they plan to use for such research, as they vary along a number of dimensions. For example, the field of view is different depending on the type of glasses used and glasses may have differing lengths of time that the technology records video and audio data. Privacy, confidentiality and

ethical issues are also present, as some users may not be comfortable with recording health-related activities and others in their immediate context or environment. This is linked to the visibility of such glasses as recording devices; for example, Google® glasses [20] were observed by others in an individual's context as recording devices, whereas Snapchat® glasses are similar to spectacles or sunglasses [15]. Additionally, advances in analysis of video data will be needed as recording using glasses could potentially lead to large data sets.

There are many challenges associated with using glasses that can audio and video record, but they do represent a significant advance in the area of mobile usability, workflow and safety testing as they allow the researcher to see how technology is used in real-world settings, capturing information about user gestures, activities and interactions with the mobile technology and their environment. As this represents a significant leap in recording devices, future research will need to focus on how to analyze user interactions between mobile devices, software and others in the healthcare setting. Such information will be critical to refining existing mobile hardware and software used by consumers for health-related activities. The work will lead to technology designs that have greater and more seamless intergration and support for healthcare activities within the context of the users's worklife and their lifestyle.

Acknowledgements

This work is supported by the Agency for Healthcare Research and Quality (AHRQ) grant R01HS023708.

References

- [1] M. Househ, E. M. Borycki, A. W. Kushniruk. mHealth: A passing fad or here to stay? In *Telemedicine and eHealth Services, Policies and Applications*. IGI Global, Hershey, Pennsylvania, 2012.
- [2] A. W. Kushniruk, M. M. Triola, E. M. Borycki, B. Stein, J. L. Kannry. Technology induced error and usability. *Int J Med Inform* **74**(7-8) (2005), 519-26.
- [3] R. Koppel, T. Wetterneck, J. L., Telles, B. T. Karsh. Workarounds to barcode medication administration systems: Their occurrences, causes and threats to patient safety. *JAMIA* **15**(4) (2008), 408-23.
- [4] E. M. Borycki, M. S. Househ, A. W. Kushniruk, C. Nohr, H. Takeda. Empowering patients: Making health information and systems safer for patient and the public. *Yearb Med Inform* **7** (2012), 56-64.
- [5] S. Fox, M. Duggan, Pew Internet. *Mobile apps*. 2012. file:///C:/Users/emb/Downloads/PIP_MobileHealth2012_FINAL.pdf
- [6] A. W. Kushniruk, E. M. Borycki *Human, social, and organizational aspects of health information systems*. IGI Global; 2008.
- [7] PEW. *Global divide on smartphone ownership*. 2016. <http://www.pewglobal.org/2016/02/22/smartphone-ownership-rates-skyrocket-in-many-emerging-economies-but-digital-divide-remains/technology-report-03-06/>
- [8] A. C. Li, J. L. Kannry, A. Kushniruk, D. Chrimes, T. G. McGinn, D. Edonyabo, D. M. Mann. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. *Int J Med Inform* **81**(11) (2012), 761-72.
- [8] T. B. Baylis, A. W. Kushniruk, E. M. Borycki, Low-cost usability testing for health information systems: Is it worth the effort? *Stud Health Technol Inform* **180** (2011), 363-7.
- [9] E. M. Borycki, H. Monkman, J. Griffith, A. W. Kushniruk. Mobile usability testing in healthcare: Methodological approaches. *Stud Health Technol Inform* **216** (2015), 338-42.
- [10] A. E. Kushniruk, E. M. Borycki, S. Kuwata, J. Kannry. Emerging approaches to usability evaluation of health information systems, *Stud Health Technol Inform* **169** (2011), 915-9.
- [11] J. Rubin, D. Chisnel, J. Spool, *Handbook of usability testing: How to plan, design, and conduct effectiveness tests* (2nd ed.). Indianapolis, Wiley.
- [12] E. M. Borycki, J. Griffith, H. Monkman, C. Reid-Haughian. Effect of the mobile phone on usability and safety of health information technology. *Stud Health Technol Inform* **234** (2017), 37-41..
- [13] L. E. Kam, W. G. Chismar. Online self-disclosure: Model for the use of internet-based technologies in collecting sensitive health information. *Int J Healthcare Tech Manag* **7**(3-4) 2015, 218-232.
- [14] O. Laurent. The doctor used Snapchat's spectacles to record a surgery. Time 2016, <http://time.com/4597434/snapchat-spectacles-snap-surgery/>
- [15] A. Balakrishnan. Why Snapchat's new glasses could be more than just a toy. <http://www.cnn.com/2016/09/26/why-snapchats-new-glasses-could-be-more-than-just-a-toy.html>
- [16] CNET. We tried Snapchat spectacles here's what it's like. <https://www.cnet.com/products/snapchat-spectacles/preview/>
- [17] M. Swider. Google glass review. Sept 27, 2016 <http://www.techradar.com/reviews/gadgets/google-glass-1152283/review>
- [18] Wikipedia. Google glass. https://en.wikipedia.org/wiki/Google_Glass
- [19] What's the best spy camera glasses? <http://www.spycameasy.com/2014/11/whats-the-best-spy-camera-glasses/>
- [20] USA Today. Pivot head glasses sport a camera for recording. <http://usatoday30.usatoday.com/tech/columnist/edwardbaig/story/2012-07-09/pivothead-glasses-review/56136656/1>
- [21] S. Mlot. Tobii glasses 2 give eye tracking a boost. May 20, 2014. <http://www.pcmag.com/article2/0,2817,2458310,00.asp>

Address for correspondence:

Elizabeth Borycki email: emb@uvic.ca

Integrated Care Processes Designed for the Future Healthcare System

Thomas Bürkle^a, Kerstin Denecke^a, Michael Lehmann^a, Erwin Zetz^b, Jürgen Holm^a

^a Institute for Medical Informatics, Bern University of Applied Sciences, Bern, Switzerland

^b GS1, Bern, Switzerland

Abstract

Previous studies showed that a stronger focus on integrated care models is required, targeting at a seamless coordinated care reaching patients even at home and thus establish patient centered and patient driven innovation activities. Such integrated care model is important among others for ensuring traceability that is necessary when problems during treatment occur or in cases of product counterfeiting for ensuring patient safety. The objective of our work is to realize an efficient, continuous care process, considering an optimal, end-to-end treatment path. In this paper, we describe the process how to pave the way for developing technologies for realizing a cross-sector treatment pathway. Following a multi-stakeholder principle and by applying requirement analysis and world café methodology we designed first concepts and strategies.

Keywords:

Healthcare; Patient Safety; Telemedicine

Introduction

“Prognoses are difficult, especially when they concern the future”, this sentence has been attributed to the Danish physicist Niels Bohr (1885-1962) in a lecture in Copenhagen. Nevertheless, prognoses for the future of hospitals, their organization and their technology have been made at all times [1]. In a recent report of the Spanish center for research in healthcare innovation management, two leading European university hospitals have been analyzed and a set of future challenges for such sites have been derived [2]. They include increasing demands combined with decreasing resources, smaller and more complex hospitals, the demand for new services such as personalized medicine and a future duplicate role providing complex care as well as acute services for the catchment area. The authors emphasize the need for knowledge driven redesign of efficient services, the requirement for distributed organizational structures and the trend towards technology driven healthcare service innovation. They foresee a stronger focus on integrated care models, a tendency towards seamless coordinated care reaching patients even at home and thus patient centered and patient driven innovation activities [2]. A strong focus on modern information technologies in hospitals is undisputed [2, 3]. In the UK, a future hospital commission has been funded by the royal college of physicians and finalized their report devoting a complete chapter to improved information technology [4, 5]. Some of the identified technology trends include connected medical devices, quantified health data,

medical records on-the-go, on-call doctors via telehealth and 3-D printing [6]. In view of limited resources, prevention and early risk detection will become a prime focus and eHealth platforms may complement today's mostly curative approaches to medicine. New forms of care will emerge, in which the patient will make self-determined decisions and will be able to self-manage his personal health.

The institute for medical informatics (I4MI) of the Bern University of Applied Sciences BFH, and GS1 Switzerland, a non-profit professional association for optimizing value added process chains cooperated in a first project “Hospital of the future” to develop a concept for innovative future healthcare based on today's available information technology [7]. Now, the second project phase has been started under the title “Hospital of the future live”. We follow the vision that information relevant for the treatment of a patient will be available to all authorized actors at the point of care in a secure, reproducible and appropriate manner. This includes not only an accurate documentation of the services and materials provided to the patients, but also improved workflows where resources and required materials will be available just in time for medical procedures.

We try to realize the vision of an integrated care model that reaches the patient already at home with a variety of projects along a typical treatment pathway. Results are implemented in our comprehensive I4MI medical informatics laboratory and in real environments to demonstrate an optimized treatment pathway and information workflow of the future. Here, we introduce the overall methodology of the project and demonstrate first results.

Methods

Both projects, “Hospital of the future” and “Hospital of the future live” have been organized as multi-stakeholder projects. The first project started in 2012 with a visionary brainstorming process focusing on the Swiss healthcare environment with the partners GS1 and *economiesuisse*, the holding organization of the Swiss economy organizations. Following the brainstorming process, an in-depth analysis was performed to determine the required medical activities for a virtual elderly and multi-morbid patient requiring hip surgery and implantation of a total endoprosthesis (TEP) for advanced arthrosis. The project maintained an advisory board with participants from hospitals, IT industry, pharmaceutical industry and medical product suppliers. Multi stakeholder table technology was employed to gain insights into the current and a possible future improved care process.

“Hospital of the future live” started in 2016 with initially 21 partners, including six Swiss hospitals, four major IT suppliers, IHE Suisse, IHE User Group Switzerland (<http://www.ihe-suisse.ch>) and eHealth Suisse, the coordinating body for the implementation of the Swiss eHealth architecture [8]. The project runs coordination meetings to define and prioritize punctual areas of activity along the patient pathway in shape of atomic work packages (WP). A total of 6 major coordination meetings with all participants have been scheduled until 2018 for this project. WPs are grouped into activities prior to hospitalization, WPs during hospitalization and WPs downstream when the patient has returned home. In the coordination meetings, the participants passed through parallel sessions in which an expert explains each WP. They had then to vote for the prioritization of the work, and commit themselves if they will participate actively in the respective WP. From this voting, we selected the most relevant WPs which in turn were condensed to projects which may comprise between one and six WPs.

These projects are further outlined and distributed to the more advanced students of the Bachelor study of medical informatics at the BFH. In shape of seminar and so called living case (practical courses) topics in which requirements from the relevant stakeholders are collected, including current processes and ideas on improved process flow, target states are specified and prototypes are implemented as basis for future product development.

Living cases will result either in a working prototype or an implemented piece of software, which is installed in the I4MI medical informatics laboratory to demonstrate a potential solution for the problem. Within seminar projects, the students have to analyze requirements in detail with the stakeholders, to develop a feasible IT concept to solve the determined interoperability problems and to describe and present this work. For the coordination of the different student activities, we use the world café methodology, which helps to identify future visions for selected processes.

Typically, a successful living case work may be continued towards the bachelor thesis which should result not only in an IT based implementation, but also in a formal evaluation of the developed solution in a real environment, e.g. a hospital department, a medical surgery, a rehabilitation center or a community care organization.

Results

First study

Results of the first study “Hospital of the future” included a document which described the patient’s current clinical pathway from home to home across several care providers and institutions comprising many process gaps and inefficiencies in today’s Swiss healthcare system and contrasted this with a visionary future clinical pathway avoiding most of these gaps and information loss [7].

The study demonstrated that Swiss hospitals are not sufficiently integrated into the entire information flow. Integration of data and processes towards the outpatient sector and the supplier is yet rarely realized. A closer look at the flow of information in the Swiss healthcare system demonstrated that, in many places, the knowledge about the patient is locked in single information systems instead of

being integrated among systems to get the whole picture on the health of a patient which is necessary when making personalized clinical decisions. The study demonstrated the need for a better linkage between medical activities along the governing clinical pathway and accompanying processes starting from the admission into a hospital to discharge, including the supply chain. Particularly those processes were closely investigated, in which the dispensing of medication and / or medical equipment or consumables were involved.

In addition, interruptions in the logistic information flow for drugs or materials themselves, which would permit establishing a continuous supply chain, have been analyzed. Such traceability is necessary when problems during treatment occur or in cases of product counterfeiting for ensuring patient safety. In many Swiss hospitals, however, the supply chain for materials is broken between the different internal storage places such as ward cupboards for medical supply or drugs. Often it is unknown, when, where and how drugs, consumables or smaller devices are ordered, consumed or stored, potentially resulting in drugs with exceeded expiry date in stock or insufficient trace back of defects in instruments or devices. This can lead to serious consequences in security and patient safety.

To achieve a better common understanding of complex information flow e.g. in the medication process, an easy to understand process tool named IXPRA (Figure 1) has been developed to support process analysis tasks in a multi stakeholder environment [9]. IXPRA stands for Interface Cross Culture Process Analysis Tool and is a toolset which describes medical processes on a functional level with main processes, sub-processes and particularities, but also on an application level resulting in clinical use cases and on an interaction level comprising involved staff, involved IT systems and special challenges. The functional level describes the basic process steps, sub-processes and comments. The application level describes the use cases or a task description with the relevant process steps. The interaction level displays involved people, IT systems and special features. Using IXPRA, the problems and challenges within a process can be identified. The results of an IXPRA process analysis form the basis for discussions with involved people and for developing new solutions that improve the processes. IXPRA enables rapid focusing on the information gaps and information loss and supports especially weakness analysis. Besides the concrete process analysis, factors such as the possible influence of management decisions and the internal incident culture of the processes can be examined. Today, an IT based implementation of IXPRA is available from MID GmbH Nuremberg, Germany.

Within the first study “Hospital of the future”, IXPRA has been repeatedly used to analyze e.g. weaknesses in the medication process at hospital partners such as the Spital Thurgau AG and Spitalzentrum Biel. We found considerable weaknesses in the investigated workflows, e.g. concerning the medication process. Currently, the information flow is broken in different dimensions as it was identified for the care process in the outpatient sector: information on the patient him/herself; information on instruments / material is collected – if at all – in separate information systems. This causes a high risk for patient safety, outcome quality and limits efficiency of workflows. In tangible terms, our goals for the second study are:

1. Realizing a cross-sector treatment pathway,
2. Integrating eHealth and demonstrate potentials of eHealth and
3. Realizing new use cases and derive business cases.

Second study

The second study “Hospital of the future live” started in June 2016 and is planned to last for two years, thus only preliminary results are available at the time of this paper. The second study builds upon the results of the first study including the detected weaknesses. So far, two of six planned multi stakeholder coordination meetings have been completed, the first one in July 2016. In this meeting, a set of 59 WPs has been jointly defined and finalized. They include:

- 18 WP prior to hospitalization. An example is the following WP: Given the situation that a patient suffers from arthrosis and also from diabetes. How can his diabetes management at home be improved using the future Swiss eHealth environment (new legislation for a countrywide electronic patient record in Switzerland)? This includes setting up a connected diabetes monitoring platform on mobile devices and establishing appropriate monitoring mechanisms to avoid hyper- and hypoglycaemic episodes.
- 21 WP center around hospital care and treatment. An example WP deals with the information flow between acute hospital care and rehabilitation care for the remobilization of the patient. The task is to examine and demonstrate which data and information can be transferred in a structured fashion across the Swiss eHealth environment between both institutions.
- 14 WP deal with the patient’s downstream home activities and the logistics linkage. It can be expected that stronger efforts will be made to support citizens to continue living in their home environment, aided by AAL (active and assisted living) activities such as an intelligent wardrobe in their home environment [10]. As an example of this type, consider a WP where we will examine the required activities and information workflows for the re-initialization of the home environment digital devices when the patient returns and may have different gait patterns which should not lead to unnecessary alerts of the fall detecting floor in his/her apartment.
- 6 remaining WPs enable us to build the required infrastructure within our medical informatics laboratory to support those activities. Within one WP, we will establish a generic middleware for connected sensor technology.

With the methods outlined before, we aggregated a first set of 28 highly prioritized WPs into 12 student projects (Figure 2).

A preliminary result concerned the preoperative handover between GP, orthopedic specialist and hospital. The current outpatient care process was analyzed with the goal to eliminate weaknesses and to derive an innovative IT-driven process. The result of the process analysis is shown in Figure 3. Several limitations could be identified in this care process:

- Referral forms are filled by hand and sent by fax.

- Reports from the orthopedic specialist are sent by fax and re-entered into software at the GP.
- Image data is transferred by the patient (manual transport of discs).
- Reports from the hospital are sent by mail.

These issues lead to information loss. Within the student project, several improvements have been identified:

First, the information system at the general practitioner (GP) could be enabled to recommend appropriate specialists, rehabilitation centers, hospitals and pharmacies, which are specialized for the medical problem of the given patient. Relevant criteria might be distance to the patient’s home location, experience in treating hip arthrosis, free capacity, and information about the available surgeon(s). This would save resources in searching for relevant specialists.

Second, using an eHealth platform with access for all relevant persons of the care team can help avoiding such limitations. Each report and examination result can be made available through the platform and authorized persons can access the data directly from there. Third, appointments with the orthopedic specialist could be made directly with IT from the GP’s place. Fourth, patients could rate the treatment for each provider through an online portal. These ratings could be made publicly available for decision making and quality assessment.

Discussion

The digital transformation is finding its way into the various national healthcare systems. E-Health strategies have been developed in the last years in multiple European countries such as Germany, Switzerland or Austria [11]. These strategies aim at improving the availability of treatment-related patient data beyond sectoral borders. But even before these goals are reached, we see the emergence of new requirements such as personalized medicine, big data, or intelligent self-monitoring devices with a yet unexploited potential to detect diseases earlier and prevent adverse events.

On the other side, all civilized nations are faced with ever increasing healthcare costs, an ageing population and the breakup of traditional family structures. We need to use the potential of modern information technology to ease these effects and to enable elder citizen to stay in their familiar home environment as long as possible.

A specific problem in the healthcare sector is that the treatment process consists of complex workflows, where many actors and IT systems are involved to achieve the most efficient and effective treatment of a patient. The processes in this workflow extend beyond the classical clinical pathway which covers the workflow within a single institution such as a hospital. Instead, it is essential to include also processes outside the hospital such as early diagnosing at the general practitioner, and collection of personal health information in the intelligent home, as well as downstream processes and the interlinked logistic activities. Benefits of such an integrated care process are manifold and can include:

- Individuals become actively involved in the treatment processes and become able to self-manage their health.

- The family medicine, pharmacy, rehabilitation or nursing services profit from new, integrated IT processes.
- Healthcare providers are provided with more support in treatment planning and better provision of pharmaceuticals and materials.
- The entire logistics chain and the participating industrial partners achieve more transparency regarding the flow of goods and consumption of goods.
- The administration can improve cost center accounting.
- The hospital as a whole benefits from optimized processes and efficiency improvements, which are an additional competitive advantage.
- The ICT industry partners are enabled by mapping future workflows that involve mobile devices and other upcoming technologies within ICT healthcare system.

We are confident that these results as well as the IXPRA method will help us to focus on promising improvements within the treatment chain. The WPs and subprojects of our efforts will result in the development of prototypical IT installations in our medical informatics laboratory, which help to understand the effect and the extent of the potential improvement which can be achieved. We can perform IT development, evaluation and verification of processes and technologies in this environment without causing harm to any person.

Obviously the next step must be a practical evaluation also in real environments and this is a declared goal for the project "Hospital of the future live" which shall give good indicators which of our innovative ideas may come up to our expectations.

Physicians and healthcare professionals may have (legitimate) reservations against new technologies workflow alterations, often in the interest of their patients. Introducing new treatment paths with support of modern IT technology requires careful action and a change management approach. We need to demonstrate and verify the potential improvement first before asking for change and our approach should enable us to do so.

Conclusion

The triggering question in this work is: How can medical informatics influence the care process and the care network positively in the next five years? We address this question by developing innovative ideas for the future healthcare processes and by showcasing them in a living lab. In this

paper, we provided an overview on a suite of projects targeting at 1) identifying limitations in processes of current healthcare systems and 2) developing solutions for designing the future healthcare system. Our goal is not only the improvement of the care process by enabling a continuous management and data integration, but also the transparent integration of the supply chain with the care process. The upcoming results of the WPs will be presented to the stakeholders and wherever possible, the results will be implemented in the real world after comprehensive testing in our labs.

References

- [1] J. Haycock, A. Stanley, N. Edwards, and R. Nicholls, Changing Hospitals, *BMJ* (1999) 319 (7219):1262-4.
- [2] J. Ribera, G. Antoja, M. Rosenmöller, and P. Borrás, Hospital of the Future, *IESE Center for Research in Healthcare Innovation Management*, Barcelona 2016, available under https://www.accenture.com/_acnmedia/PDF-7/Accenture-Hospital-of-the-Future.pdf
- [3] P. Attfield, The Hospital of the future, *The Globe and Mail* May 2nd, 2016, available under <http://www.theglobeandmail.com/report-on-business/industry-news/property-report/the-hospital-of-the-future/article29823293/>
- [4] T. Evans. Future Hospital Commission: the time for action. *Br J Gen Pract* (2013) Nov;63 (616):571-2.
- [5] The Future Hospital Commission. Future hospital: Caring for medical patients. Royal College of Physicians 2013. ISBN 9781860165184
- [6] J. White. 5 technology trends that will affect hospital care, *Healthcare Business & Technology* Aug 6th, 2015, available under <http://www.healthcarebusinesstech.com/technology-future-hospitals/>
- [7] Spital der Zukunft (Hospital of the future) <https://www.gs1.ch/gsl-system/gesundheitswesen/spital-der-zukunft> (last access: 17.12.2016)
- [8] C. Lovis, A. Schmid, and S. Wyss, ehealth Suisse – coordinating e-health in Switzerland, *Healthcare IT Management* 6(3) (2011): 46-47.
- [9] GS1 and Bern University of Applied Sciences, Interface Crossculture Process Analysis Tool. <http://ixpra.ch/>
- [10] P. Schaad, S. Basler, M. Medini, I. Wissler, T. Bürkle, and M. Lehmann, The "intelligent wardrobe", *Stud Health Technol Inform* (2016) 225:213-7.
- [11] KA. Stroetmann, J. Artmann, and VN. Stroetmann, European countries on their journey towards national eHealth infrastructures (Final European progress report), January 2011, http://ehealth-strategies.eu/report/eHealth_Strategies_Final_Report_Web.pdf (last access: 17.12.2016)

Address for correspondence

Prof. Dr. Thomas Bürkle

Bern University of Applied Sciences

Quellgasse 21, 2502 Biel, Switzerland

thomas.buerkle@bfh.ch

The diagram illustrates the process of knowledge engineering, divided into two main sections: a high-level process and a detailed functional flow.

High-level Process: A sequence of three circles connected by arrows: **Analyse** (blue), **Understand** (yellow), and **Improve** (orange).

Functional Flow: Three levels of knowledge engineering are shown, connected by arrows:

- Functional Level (Blue):** Contains **Medication** and **Incomplete Data**.
- Application Level (Green):** Contains **Medication Documentation**.
- Interaction Level (Orange):** Contains icons for **MEDS**, **RMO**, **PROVIDER**, and **PATIENT**.

To the right of the Interaction Level, four icons represent the domains of knowledge engineering: **PEOPLE** (person icon), **IT** (car icon), **PROBLEMS** (warning triangle icon), and **SOLUTIONS** (person with speech bubble icon).

Actors

GP, pharmacy, nursing service | GP, patient | Specialist | GP, specialist | Physician, supplier, specialist | Physician, supplier, pharmacy | Physician | GP, pharmacy, nursing service

Location

Home | GP | Specialist | Pre-operation | Operating room | Post-operation | Reha | Home

Activities

Home care, monitoring | Anamnesis, examination, diagnostics | Examination, diagnostics, indication | Additional examinations, preparation of surgery | Surgery | Hospital care, treatment, monitoring | Training | Home care, monitoring through ambient systems

Work packages

Diabetes management | Prescription app | Counseling in the pharmacy | Rental of medical devices | Selection and management of clinical path | Clarification and planning hip TEP | Planning of a surgery | Patient experience | Teleprocurement (Order-to-Cash) | Material (Master Patient Data) management | Medication process in the hospital including conversion | Discharge app | Counseling pharmacy | Advice

Data flow

Patient recognizes symptoms | Diagnostic, transmission | Additional diagnostic, GP-referral | Upcoming examinations | Progress documentation, surgical report | Discharge summary, further procedures | Discharge planning, additional procedures, rehab | Feedback

```

graph TD
    subgraph General_practitioner [General practitioner]
        A[Patient admission] --> B[Physical examination including Xp]
        B --> C{Fractured? (no effusion?)}
        C -- No --> D[Additional diagnosis  
fracture or sprain/strain or  
dislocation]
        C -- Yes --> E{Selection of  
imaging modality}
        E --> F{Advanced?}
        F --> G{Yes}
        F --> H{No}
        G --> I{Referral to  
orthopedic specialist}
        H --> D
    end

    subgraph Orthopedic_specialist [Orthopedic specialist]
        I --> J[Repetition of form]
        J --> K[Consultation,  
additional examinations,  
therapy indicated]
        K --> L{ }
        L -- No --> M[Additional  
therapy actions]
        L -- Yes --> N[Admission]
        M --> O[Discharge - early  
medication? (to ERM)]
        O --> P[Discharge - late  
medication? (to ERM)]
    end

    subgraph Hospital [Hospital]
        N --> Q[Surgery]
        Q --> R[Hospital stay]
        R --> S[Discharge]
    end

    D --> T[Referral to  
orthopedic specialist]
    T --> I
    T --> U[Referral form  
(open access)]
    T --> V[Image file on CD]
    
```

The flowchart illustrates the clinical pathway for a patient with a suspected fracture of the distal radius/ulna, divided into three main sections: General practitioner, Orthopedic specialist, and Hospital.

General practitioner section:

- Patient admission** leads to **Physical examination including Xp**.
- A decision point **Fractured? (no effusion?)** follows.
 - If **No**, the pathway leads to **Additional diagnosis: fracture or sprain/strain or dislocation**.
 - If **Yes**, the pathway leads to **Selection of imaging modality**.
- From **Selection of imaging modality**, a decision is made on whether the case is **Advanced?**.
 - If **Yes**, the pathway leads to **Referral to orthopedic specialist**.
 - If **No**, the pathway leads back to **Additional diagnosis: fracture or sprain/strain or dislocation**.

Orthopedic specialist section:

- The **Referral to orthopedic specialist** leads to **Repetition of form**.
- This leads to **Consultation, additional examinations, therapy indicated**.
- A decision point follows:
 - If **No**, the pathway leads to **Additional therapy actions**.
 - If **Yes**, the pathway leads to **Admission**.
- From **Additional therapy actions**, a decision is made on whether **Discharge - early medication? (to ERM)**.
 - If **Yes**, the pathway leads to **Discharge - late medication? (to ERM)**.
 - If **No**, the pathway leads to **Referral to orthopedic specialist**.

Hospital section:

- Admission** leads to **Surgery**.
- Surgery** leads to **Hospital stay**.
- Hospital stay** leads to **Discharge**.

Additional information:

- A **Referral form (open access)** is associated with the referral to the orthopedic specialist.
- An **Image file on CD** is associated with the referral to the orthopedic specialist.
- A **Referral to orthopedic specialist** is associated with the referral to the orthopedic specialist.
- A **Referral to orthopedic specialist** is associated with the referral to the orthopedic specialist.

Figure 3 – Diagnosis process for patients with hip arthrosis in the ambulatory sector. Data transferred are shown in red. The general practitioner, an orthopedic specialist and the hospital are involved in this process.

Design and Evaluation on the Mobile Application of Transcutaneous Electrical Nerve Stimulation (TENS)

Ching-Lung Cheng, Li-Hui Lee, Yu-Ting Cheng

Department of Health Care Management, National Taipei University of Nursing and Health Science, Taipei City, Taiwan, China

Abstract

This study aims to design a transcutaneous electrical nerve stimulation Application (TENS App) according to the suggestions from potential users. To the best of our knowledge, this is the first App including meridian and acupoints for TENS. After its development, there are eight participants recruited for evaluating the usability. Despite two out of eight users reporting that the typical TENS system requires lower cost and has better functionality than TENS App, the results show that almost seventy percent of participants have a better perception of TENS App on price, functionality, convenience, operational ability, and quality. However, participants still reported concerns about the safety issue of adopting TENS App. Therefore, for people who are the first time or unfamiliar with TENS App, instructions from occupational or physical therapists are recommended. We conclude that by using TENS App, users can not only use the portable electrotherapy devices at anyplace, but also reduce their outpatient visits.

Keywords:

Electric Stimulation Therapy; Smartphone; Mobile Applications

Introduction

Pain is the unpleasant feeling induced by illness, physical impairments, or harmful outer stimuli. The two major causes of pain include the activation of stimuli to nociceptor and impairment of human nervous system. Nociceptor receives chemical, thermal, or forceful stimuli which damage human tissues and transmits painful sensation to the nervous system. And if the nervous system is impaired due to injury or illness, it will induce neurological pain or neuropathy.

Past related studies found evidence to support short-term physical therapy intervention to be effective in relieving knee joint pain or joint arthritis pain. It is also recommended to provide intensive treatment program using transcutaneous electrical nerve stimulation (TENS) with pulse rate setting between 1Hz to 150 Hz, a minimum of 20 minutes as treatment duration each session, for a minimum of 5 sessions, in a period of two to four weeks[1-3].

Clinical physical therapy relieves pain based on the Gate Control Theory (Figure 1). Based on this theory, TENS device utilizes subtle low-frequency electronic current to stimulate epidermal nerves [4]. The signals from the device are transmitted faster than pain signals, which arrive at the "gate" first and block the pain signals from passing through, thus reduce the pain perceived by the brain. TENS has been extensively used by physicians and therapists in physical medicine and rehabilitation (PM&R), and most of the TENS devices apply electrical stimulation to skin tissues, using low-

frequency pulse rate changes to induce the release of pain relief substances from central nervous system.

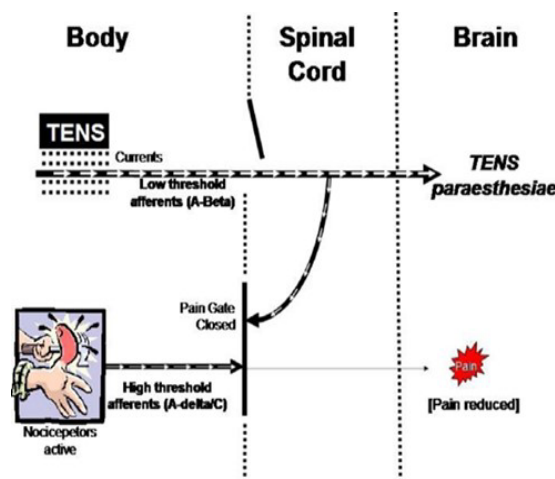


Figure 1 – Gate Control Theory.

TENS for pain relief requires customization according to the characteristics of pain, and it requires regular intervention. Physical therapy intervention with TENS devices involves with a professional assessment and adjustment protocols. Patients usually can only receive the services in clinics or hospitals. Despite its effectiveness to relieve pain, the accessibility of TENS is insufficient. Therefore, some companies such as Panasonic, Koka, Tatung, and Omron have been selling similar products for patients to use at home. Most of the devices can be operated through the equipment interfaces but few devices can be handled or configured by App, such as iTENS [5]. Moreover, these few Apps do not show the meridian and acupoints to precisely guide users in the use of pain patches. Also, there was a noticeable absence of research projects dealing with the usability of Apps for TENS.

Moreover, the portable electrotherapy pain relief patches using the latest technology have appeared on the market. The new product applies professional TENS technology to a thin pad design. It is compact in size and reusable with power saving technology, providing users more convenience and reducing waste from traditional disposable pain patches. This also presents an opportunity to design a new health information technology with portable TENS. This study aims to design a TENS App including meridian and acupoints information and then to assess its usability in five dimensions (price, functionality, convenience, operational ability, and quality) and do the comparison between the typical TENS device and our TENS App with summarizing their advantages and disadvantages.

Literature Review

This study reviews articles regarding the medical concepts behind TENS, myofascial pain, and related physical therapy intervention, as well as the clinical application of TENS systems. It aims to summarize the studies and understanding on TENS systems.

Medical Concepts behind TENS

TENS product has been widely used for palliative care in developed countries, including the non-noxious or acute pain from malignant neoplasms or related medical treatments. The product is safe, low in cost, and free from serious side effects. Under the directions of a physical therapist, people with muscular pain or patient with needs could operate the product at home and receive short-term pain relief effects. Therefore patients could apply two self-adhesive patches on their skin, and use the low-frequency electrical stimulation with control of set protocol safely [1].

Literature reviews found that the first records of using electrical stimulation to relieve pain date back to a Roman physician [2]. In 1786, the Roman physician experimented on the frog leg and proved electrical stimulation can relieve pain and reduce muscle tension. The observation along with the progress of power generation further urges the use of electrical stimulation to treat various diseases and relieve pain. In 1965, Doctor Ronald Melzack from Monterey, Canada and Doctor Patrick Wall from England published the focus treatment mechanism in University of London. Their mechanism is that the central neuron system modifies the quality and quantity of neuron signal from the peripheral neuron system to the brain[3]. Based on the Gate Control Theory, they proposed that by selectively utilizing low-frequency electrical stimulation to large diameter nerve through skin tissues, transmission of pain signals could be inhibited and thus reduce pain perception.

TENS devices are usually made of a single line (two electrodes) or two lines (four electrodes). There are potentiometer or regulator on the devices for the users to adjust the stimulation intensity. The stimulation variables include wave formation, frequency, and intensity. In sum, when utilizing TENS devices, chronic patients can choose protocols with low frequency and high-intensity stimulation, while acute patients are recommended to use protocols with high frequency and high-intensity stimulation.

Common myofascial pain and related physical therapy intervention

Myofascial Pain Syndrome is not a specific disease. Instead, it is a syndrome caused by extensive muscle tension. If the muscle is under tension status for a long period, it will impede peripheral blood flow and thus impedes the circulation system to take away waste productions from metabolism mechanism such as lactic acid. The accumulation of lactic acid in specific parts would then induce soreness and discomfort. Therefore, Myofascial Pain Syndrome is one of the common causes of low back pain.

If human body keeps in the same position for a long period, there must be a group of muscles stay in tension status for a long time. As time goes by, Myofascial Pain Syndrome occurs with muscle soreness and fatigues. Different from medical diagnosis by physicians, physical therapist evaluates and assesses patients using structure diagnosis, providing a diagnosis of somatic dysfunction and setting intervention plans accordingly. In general, physical therapist diagnoses patients under the three principles A-R-T:

- Asymmetry (A): Physical Therapist assesses patient's symmetry on the musculoskeletal system through palpation and observation.
- The Range of Motion (R): The assessment focuses on the range of the movement and its movement quality of a single or multiple joints, such as hyper- or hypomobility. The patient follows directions to check the active and passive range of motion at different joints.
- Soft Tissue (T): Physical Therapist evaluates abnormal soft tissue textures, including skin, muscle, fascia, tendon, etc.

Clinical Application of TENS System

The clinical experiments had proved a regulation rule for pain treatment. The Randomized Controlled Trials (RCTs) in the early period indicated that many patients used TENS to treat focal pain and acute pain. Carroll found that TENS did not relieve pain after surgery because no significant difference was found between the experiment group and control group on pain relief [4]. Researchers suggested that TENS could relieve chronic pain. Although this was a commonly accepted concept, current studies on TENS are not extensively enough to provide sufficient evidence on its effectiveness for rheumatoid arthritis, neck disorders, or chronic low back pain [6]. On contrast, several RCTs have provided evidence for TENS on treating pain symptoms for knee joint arthritis [7] and chronic skeletal muscle pain [8]. Some researchers report that TENS can improve pain symptoms for some patients with chronic disorders including focal muscle pain, postherpetic neuralgia, trigeminal neuralgia, phantom limb pain and diabetic neuropathy. Therefore, applying TENS to muscle inflammation on ipsilateral or contralateral parts is likely to reduce the chronic pain response over bilateral limbs caused by unilateral inflammation [9].

Methods and Study Subjects

TENS App Mechanism

If blood flow and massage could alleviate the pain, it is seen as the effect of relaxing muscle pain. This App could transport the electric energy from the device to muscle via the audio card in the smartphone. Users could easily attach two electrode pads on their skin surface to conduct energy. The electrotherapy can alleviate the pain and relax the strained muscle through the muscle layer. Then the power improves the blood circulation to decrease pain substance in the body. Both mechanisms have the pain alleviating effect.

Different low-frequency current has a different effect on alleviating and relieving pain. The effect is caused by activating neural functionality and the power of body fluid as well as increasing blood flow in the body to gradually decrease the muscle hypertonia. TENS App provides multiple functionalities which can easily control the current mode, time duration, and intensity. Users can configure its functionalities to active the effects of rhythmic muscle contraction on unbalanced muscles.

TENS App Design

Authors make many principles for the interface design to maximize the operational efficiency. The design principles include the visualization, feedback, response, synchronizing, and responsibility. We could improve and raise convenience in the TENS equipment through replacing the traditional TENS

devices by the interface on the smartphone. The study is based on the interview with a focus group [10]. After TENS App is developed. We invite those study subjects who have at least one year experience of using other TENS system to join the study. We have an interview with the study subjects for 60 to 90 minutes in a semi-constructional way. Then we can summarize the features for the prototype interface which can fulfill with designer's requirement. Furthermore, we can find out its advantages and disadvantages.

Method and Subjects

The study subjects who finished the questionnaire were aged between 20 and 50, with myofascial pain syndrome on upper back and neck, and pain lasting for at least one week. We excluded patients have operation history on the upper back, neck or vertebra. We invited 10 subjects for the interview through purposive sampling [11; 12], eight respondents finished the interview. Participants were contacted in order to explain our research purpose. After we had their agreement, we had an interview by questionnaire with every single one to listen to their experience with using the TENS App. All participants had taken the typical TENS (X brand) therapy and TENS App self-relax therapy for four times within two weeks. All participants were evaluated by a five-dimension questionnaire which is modified according to Liu's [13]. These five dimensions are in regard with users' opinion on the price, functionality, convenience, operational ability, and quality. We used their feedback about these two devices to evaluate their effects.

Results

TENS App Design

The App primary functionality includes a brief instruction film about how to use the system, a Chinese medicine acupuncture point and symptom chart, a symptom and acupuncture chart for office workers, a symptom and acupuncture chart for students, male common symptom and acupuncture point chart, female common symptom and acupuncture point chart, Chinese medicine meridian charts and reflection area lists. The homepage combines all symptoms with acupuncture points based on Chinese medicine theory. It could help users understand the relationship between the symptoms and acupuncture points (Figure 2).

We upload the instruction film on the Internet for the purpose of publicizing and advocating this product to more people and providing users an easy know-how for using this product. The built-in homepage in App also helps the first time users and individuals who are interested to know more about this App and detail (Figure 3).

The system is designed to activate the TENS operation mode automatically when the earphone connects to the smartphone. The purpose of TENS interface in this system is to provide an appropriate stimulus wave, frequency, duration and power export. Users can undergo their physical relaxing therapy at home based on their status and physiatrist or physical therapist's instructions. The duration in the system is between 10 to 60 minutes. There is a total of three different modes for users in the electrotherapy setting mode. The operation interface provides 13 levels of energy (Figure 4).



Figure 2 – The Visualization of Symptoms And Acupuncture Points.



Figure 3 – Providing Users a Know-How for Using TENS App.



Figure 4 – Different Currency Mode, Time Duration, and Intensity Selections.

Also, the built-in main symptom and acupuncture points charts show the users where the correct points are to relieve the pain on the body (Figure 5). The terminal of this device has two electrode pads attached to the human skin to relax the pain in muscle layers by electrotherapy (Figure 6). Its convenience lies in that it can increase the patient's willingness to do the portable electrotherapy at home.



Figure 5 – Acupuncture Points Charts Show the Users where the Correct Points are to Relieve the Pain on Body.

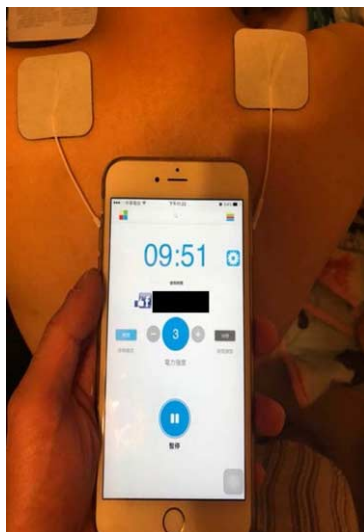


Figure 6 – The App with Two Electrode Pads Attached to Human Skin.

Demographics

In total eight respondents used similar typical TENS devices and responded to the questionnaire. The basic information is presented in Table 1. There were six female and two male respondents aged between 28 and 50. Their career sectors included service, catering, public affairs, and construction industry. There were four managers and four office administrators.

Table 1 – The Basic Information About the Study Subjects.

Respondents	Age	Gender	Occupation	Industry
A	28	Male	Office Administrator	Service
B	32	Female	Secretary	Service
C	35	Female	Accountant	Service
D	44	Female	Middle-class Manager	Public Affair
E	47	Female	General Administrator	Construction
F	48	Male	Middle-class Manager	Service
G	48	Female	Secretary	Carting
H	50	Female	Senior Manager	Carting

Assessment

This study used the X brand a conventional TENS device on the market and TENS App (named as App) to do an assessment analysis. The study endpoints in the evaluation are in Table 2. Almost seventy percent of respondents (67.5%) have a better perception of price, function, convenience, operation, and quality in operating TENS App, even two of respondents believed that the X brand device has better price and functionality. There are some suggestions from users such as the professional assistance or instruction from the occupational therapist or physical therapist to operate TENS App for the first time user. The user should use TENS App according to their pain status to alleviate the pain correctly and safely.

Table 2 – The Summary for the Advantage and Disadvantage from Respondents.

	Price		Functionality		Convenience		Operational ability		Quality	
	Brand	X	App	X	App	X	App	X	App	X
Respondents	A		V		V		V		V	
	B		V		V		V		V	
	C				V		V		V	
	D		V		V		V		V	
	E		V		V		V		V	
	F	V			V					V
	G		V				V			V
	H		V		V				V	V
count		1	6	3	6	0	8	2	6	2
better		1	6	1	4	0	8	0	4	0

Discussion

We only took consideration on Asian user's experience from eight study subjects with their agreement. All potential users' feeling can not be generalized. This study could help TENS App's upgrade and development in the future. So far we could not figure out the issue about the automatic detection of human acupuncture points. We suggest that it could compare our result with the big data to understand more about the relationship between the currency frequency and time duration. That information could help users improve and relax their body even faster. Moreover, we suggest that users could play music with the therapy for relaxing and balancing their mind.

Conclusions

Based on the comparison between the typical TENS device and TENS App, we find that participants are more focused on quality and convenience. TENS App has much better performance on these end points than the conventional ones. TENS App also has more advantages on the price than the regular ones.

To our knowledge, this is the first TENS App that was installed on a smartphone with the Chinese medicine acupuncture point charts. This helps users to find the correct position immediately that could help relax the strained muscles and relieve the pressure quickly. Furthermore, because of the high use rate of the smartphone around the world, it is possible to advocate and advertise the use of TENS App widely.

References

- [1] J.M. Bjordal, M.I. Johnson, R. AB Lopes-Martins, B. Bogen, R. Chow, and A.E. Ljunggren, Short-term efficacy of physical interventions in osteoarthritic knee pain. A systematic review and meta-analysis of randomised placebo-controlled trials, *Bmc Musculoskeletal Disorders* **8** (2007).
- [2] P.L. Gildenberg, History of electrical neuromodulation for chronic pain, *Pain Medicine* **7** (2006), S7-S13.
- [3] R. Melzack and P.D. Wall, Pain mechanisms: A new theory (Reprinted from Science, vol 150, pg 971-979, 1965), *Pain Forum* **5** (1996), 3-11.
- [4] D. Carroll, M. Tramer, H. McQuay, B. Nye, and A. Moore, Randomization is important in studies with pain outcomes: Systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain, *British Journal of Anaesthesia* **77** (1996), 798-803.
- [5] iTENS, iTENS APP, in, 2017.
- [6] C.L. Hsieh, J.G. Lin, T.C. Li, and Q.Y. Chang, Changes of pulse rate and skin temperature evoked by electroacupuncture stimulation with different frequency on both Zusanli acupoints in humans, *American Journal of Chinese Medicine* **27** (1999), 11-18.
- [7] L. Brosseau, M.G. Judd, S. Marchand, V.A. Robinson, P. Tugwell, G. Wells, and K. Yonge, Transcutaneous electrical nerve stimulation (TENS) for the treatment of rheumatoid arthritis in the hand, *Cochrane Database Syst Rev* (2003), CD004377.
- [8] P. Kroeling, A.R. Gross, C.H. Goldsmith, and G. Cervical Overview, A Cochrane review of electrotherapy for mechanical neck disorders, *Spine (Phila Pa 1976)* **30** (2005), E641-648.
- [9] K.A. Sluka and D. Walsh, Transcutaneous electrical nerve stimulation: Basic science mechanisms and clinical effectiveness, *Journal of Pain* **4** (2003), 109-121.
- [10] 鄭夙芬, 焦點團體研究法的理論與應用, *選舉研究* **12** (2005), 211-239.
- [11] K.M.T. Collins, A.J. Onwuegbuzie, and Q.G. Jiao, A Mixed Methods Investigation of Mixed Methods Sampling Designs in Social and Health Science Research, *Journal of Mixed Methods Research* **1** (2007), 267-294.
- [12] L.J. Suen, H.M. Huang, and H.H. Lee, A comparison of convenience sampling and purposive sampling, *Hu Li Za Zhi* **61** (2014), 105-111.
- [13] 劉智凱, 行動化音樂紓壓聆聽系統, 中國醫藥大學, 2014.

Address for correspondence

Li-Hui Lee

Mailing address: No.89, Neijiang St., Wanhua Dist., Taipei City 10845, Taiwan

E-mail: cclhlee@ntunhs.edu.tw.

Phone number: +8862-23885111#6126.

In Search of a Digital Health Compass to Navigate the Health System

Catherine Chronaki^a, Robert Stegwee^{b,c}, Anne Moen^d

^aHL7 Foundation, Brussels, Belgium, ^bCGI, ^cCEN/TC 251, The Netherlands, ^dUniversity of Oslo, Norway

Abstract

Healthcare systems increasingly rely on digital technologies to sustain costs and improve access to quality care. Data drive a wave of automation aspiring to improve productivity by forging connections between health and wellness, medical research, and clinical decision support. Mobile apps and patient-generated data combined with provider recordings pave the way towards personalized care pathways and just-in time access to health services. Navigating the health ecosystem becomes challenging as roles and relationships change. This paper reflects on the digital health compass to navigate the health system using one's own data. Health information technology standards are at the core of the compass, to tap the potential of shared aggregate data and sustain trust. The notion of the patient summary as a window to one's health is used as an example to drive our call for action for health informatics to develop methods to calibrate the digital health compass and feed on 'my data', respecting 'my decision', to fuel 'our ePower'.

Keywords:

Telemedicine; Reference Standards; Personal Health Records

Introduction

Knowledge is power. Despite extensive investments in digital health technology and incentives for uptake of eHealth consumer services, navigating the health system online is hard. The 2014 digital health literacy survey confirms differences and widening divides [1]. The "Inverse Care Law" proposed by Hart in 1971 [2], seems to apply to eHealth. Availability of, access to, and productive use of advanced medical or social care services and digital health tools, varies inversely with the needs of the communities. In fact, the low adoption of digital health technology and eHealth services among segments of the population underscores persistent disparities in health care.

Barriers and challenges are not to be underestimated as people feel that with excessive use of technology we may lose the human touch. Taking into account culture, education, skills, costs, perceptions of power and role, is essential for successful community actions. These elements come together in digital health literacy, "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to address or solve a health problem" [3], which underpins the knowledge and skills required to construct the digital health compass to navigate the healthcare system. Patients living with an implanted device, coping with chronic disease, engaging in self-care, caring for an elderly relative in deteriorating health, or an ailing child, need a digital health compass. Hyper-personalization enters every facet of our life

and personal data on health and behaviors can be used to present information and knowledge, services, and tools, at the right time and in the right way [4].

There is an ongoing shift in the terrain of the health sector from cure to care, where patients, families and informal care givers actively participate. This may be because they want to help [5] or they are expected to make health decisions [6, 7]. We argue that citizens need a digital health compass to unlock the power of their health information, drive their engagement in personal health information management, and identify the most suitable health services for the situation at hand. With the increasing quantity and uneven quality of health data, patient summaries could be the starting point to write an individual's health story.

Patient summaries can point to key information accumulated across health systems, sites and care settings and essential information in planned and emergency encounters. However, this assumes a joint effort to clarify expectations and keep patient summaries accurate and complete [7,8]. The Trillium Bridge project recommended a patient summary standard for people to access and share their health information and to drive incentives for quality record keeping and health professional appraisals¹.

Attending to data provenance, patient summaries complemented with patient-generated data can supplement health services and facilitate a smooth transition to digital health. High quality patient summaries can foster safe and effective care *for or by* an individual in a variety of situations. For example, we may offer our patient summary during an emergency hospital admission. We may use them to seek a second opinion or search health information online, or just to connect health professionals that typically do not exchange information or cooperate.

Experienced users may employ patient summaries in connection to personal health records to monitor their health and engage in health decisions. They may collect personal observations to complement or expand on existing data, to support self-care and follow up on health management activities. A data culture would take us further, to where we use patient summaries to safely engage in personalized navigation of the internet and network with people facing similar health issues.

Progress however, is slow. Reasons can be summed up with low digital health literacy. For the most part, patients and informal caregivers are not excited by available digital health tools. Either they do not trust them, or do not know how to use them. To make matters worse, many health professionals are not comfortable with recommending specific digital health tools to their patients. Limited sharing of information and under-developed cooperation among patients, informal care givers and health care professionals leaves untapped the potential of data for informed health decisions. Resistance is fueled by the need to protect our privacy, dignity, integrity, and individuality. Where

¹http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=11039

the Internet of Things (IoT) meets health care, a plethora of tools, gadgets, and apps overpromise and under-deliver on improving health and wellness and supporting an active lifestyle. They score low in actionable knowledge, partly because disconnected from health and social care services, they hinder personal efforts to share intimate information, while preserving one's self-reliance, autonomy, and freedom of choice.

The rest of this paper is organized as follows. The next section presents the vision of the digital health compass and its relation to health information interoperability standards that link fragmented sources of information. The case of the patient summary illustrates how the compass may work from the perspectives of health systems, eHealth consumers, the healthcare workforce, and the eHealth market (see Figure 1). Then, "my data", "my decision", "our ePower" complete the vision of the digital health compass with a call for action to the medical informatics community that can help shift the narrative towards safe and trusted use of health data to benefit individuals and the society as a whole.



Figure 1– A Digital Health Compass to Navigate the Health System.

Digital health compass

A digital health compass with knowledge of a person's digital health literacy profile can point to eHealth resources that foster personal control and empowerment. Different dimensions need to be considered to understand how a digital health compass can support safety, prevent harmful events, and assist in managing efficient, connected services of high quality and relevance in the digital health ecosystem. Health data standards, open interfaces, and a culture of sharing increase trust. Complementary initiatives to health information technology standards are the Dublin Core Metadata Initiative (DCMI) headings, the HONCode labelling online health resources, and W3C guidelines for usability and accessibility building confidence in navigation.

Standards in the digital health ecosystem

Health information technology standards are required to provide common metadata about digital health products and assemble fragmented information scaling up and sustaining digital health literacy [5, 9]. Standards developing organizations work together on standards to meet the health information needs of people within and across health facilities. The value of data and the increasing focus on patient experience, dictates global cooperation on open standards emphasizing mobile use.

HL7 FHIR

In HL7, this trend is reflected in Fast Healthcare Interoperability Resources (FHIR), a new standards initiative seeking to liberate data for population health and precision medicine in the context of learning health systems. FHIR is based on a set of

modular components or *Resources*, which are small discrete units of exchange with defined behaviour addressable using URLs. Resources are combined into profiles to solve practical clinical and administrative problems. *Resource extensions* and *profiles* facilitate the addition of data that are not part of the core, which follows the 80/20 rule i.e. cover the top 80% use cases and most frequent functionality. The overall approach is web-based (i.e. RESTful API), service driven, and supports adaptation and portability of components moving program code with data as *resource bundles*. A robust FHIR resource *maturity framework* tracks the stability and extent of world-wide adoption of every resource. The version of FHIR balloted in 2015 (DSTU2), includes 27 clinical resources along with supporting financial, conformance, workflow, identification, and infrastructure resources. The high appeal of FHIR can be attributed to its vibrant and committed implementation community, available tools, services, and data to learn, explore, and experiment with minimal cost in time or money. ClinFHIR is such a tool specifically designed for clinicians². With more integrated and specific systems and more information available in real-time, "live" sharing of health data becomes the new norm, and FHIR resources are the ideal vehicle for digital health literacy. It is these characteristics of HL7 FHIR that allow it to function as infrastructure for interoperability and innovation, by connecting the dots from user experience to data exchange on the wire.

Standards for consent management [10] support the aim to increase security, trustworthiness and transparency of digital health services making individuals and organizations comfortable with sharing or donating data to the community. ISO/TC215, CEN/TC251, and HL7 with active participation from regulators like EMA and FDA, work on standards for the identification of medicines. The openMedicine³ project focused on the identification of medicinal products throughout their lifecycle with particular emphasis on cross-border ePrescription and eDispensation. Combining this information with the personal health data of the patient fuels innovation. For example, one can imagine strolling into a Pharmacy and consulting an app on over the counter medications most appropriate for them. We may also consider active personalized medication leaflets that adapt to their health and lifestyle offering alerts and enhancing medication compliance. Essential to the digital health compass is the use of reference clinical terminologies that contribute to higher data quality. SNOMED CT is active in this area with consistent efforts in user-interface terminologies to pave the way towards individualized digital health services.

Health on the Net code of Conduct

The Health on the Net code of conduct or HONcode [11] is a process metric to determine if the construction and maintenance of a website conform to approved standards of excellence. The HONcode is used to certify health websites with content respecting the HONcode criteria (see Figure 2). Compliance to the HONcode of conduct is reflected by displaying the HONcode seal. The HONcode contributes to the digital health compass by helping individuals make informed choices in accessing trustworthy content. Extensions of HONcode for health and wellness apps can be envisioned as a natural next step.

² <https://fhirblog.com/2016/11/06/clinfhir-profiling-walk-through/>

³ www.openMedicine.eu and www.assessCT.eu

Principle	Detail
HC1 Authority	Indicates the qualifications of the authors
HC2 Complementarity	Information supports, does not replace, the doctor-patient relationship
HC3 Privacy policy	Respects the privacy and confidentiality of personal data submitted to the site by the visitor
HC4 Attribution of - reference criteria - date	Cites the source(s) of published information Dates medical and health pages
HC5 Justifiability	Backs up claims relating to benefits and performance
HC6 Transparency	Presentation is accessible; contact information is present
HC7 Financial disclosure	Identifies funding sources
HC8 Advertising policy	Clearly distinguishes advertising from editorial content

Figure 2—Principles Underlying the HONcode of Conduct.

Health information technology standards assemble fragmented sources of health information, nurturing trust. Answering how different standards fit together and how joint products are governed and adapted to the changing needs of the community is essential. This is expressed in the vision of the eStandards project⁴ imagining: “...a global eHealth ecosystem where people receive timely safe and informed health care, anywhere around the globe and interoperability assets fuel creativity, entrepreneurship, and innovation, as digital health standards nurture large-scale eHealth deployments and enable co-creation with trusted dialogs on costs and plans that drive great expectations.” eStandards cooperates with the health informatics community on a roadmap for cooperative standards development, to form the core of the digital health compass.

The case of patient summaries

Patient summary initiatives to make health information available to patients and families are supported by the EU Directive 2016/679/EU on Data Portability. The directive states the right of individuals to transfer their personal data from one electronic processing system to another, in a machine readable format. Combined with Directive 2011/14/EU on patients’ rights to crossborder healthcare and Directive 2013/37/EU on re-use of public sector information, patient summaries can be rethought as a tool for digital health innovation.

As part of the digital health compass, the patient summary helps individuals use their data and collect observations to develop the knowledge and confidence to select digital health tools that are right for them and engage productively in the digital transformation of the society. Patients, families and informal care givers join forces with the health team supported by research and development policies that advance digital health literacy. Health informatics supported by information technology standards and open application interfaces can serve as catalysts in the continuous improvement of data quality and the nurturing of trust relations in networks that transcend organizations, health systems and countries.

Trillium Bridge recommended that the patient summary includes at a minimum problems and procedures, medications and implantable devices, vaccinations, and allergies. Additional data to be included are labs, diagnostic images, and encounters. With FHIR resources one may retrieve specific sections of the patient summary or combine them in a clinical document. The low initial cost of working with FHIR opening up data sources traditionally behind organizational walls, helps shape the data sharing culture essential to construct the digital health compass.

Patient summaries can feed dashboards, the starting point in the search of health professionals for more detailed health information about a patient. Patient generated data can be summarized in this dashboard to assess progress towards achieving mutually agreed health goals supported by digital health literacy interventions. On the community level, patient summaries could help keep track of the health needs of the population and become an indispensable tool for evidence-based policy.

For the digital health compass, co-creation of patient summaries by patients, health professionals, and informal care givers comprises elements of digital health literacy, trust, and service provision. The key element of digital health literacy is understanding of health information for care, wellness, prevention, and engagement. To co-create in trust, data integrity of contributions from patients and providers and stewardship when sharing, interpreting, and complementing health information are essential. Finally, an element of service provision rooted in health experience and team play, advocates to liberate the data in the name of innovation and progress.



Figure 3—Dimensions of the Digital Health Compass.

Personal experience is key in understanding the perceived impact on a person’s life, disease progression and expected outcome and impact of treatment options for individual patients in relation to their lifestyle choices. Tools for capturing, analysing, and relating all this data are becoming available with the person at the centre! First hand experts are the levers to adoption and they are not by definition technology driven. Frequently they are suffering from cognitive impairments and mutual trust is essential for them to willingly share the effects of the disease on their daily life and for technology experts to meet their needs. They are the ones who can tell us the real story behind the effects of this disease and its diagnosis, from their own experience with the disease. However, although patient needs should drive demand, frequently patients are in the back seat, while sustainability of the health system, shortages in the workforce, fear of the unknown market regulation and traditional roles drive. Thus, we need to examine the perspective of health systems, workforce, consumers, and market guided by the core dimensions of the digital health compass: educate, enact, evaluate, empower, evolve, enable, and empower, shown in Figure 3.

Perspective of health systems

The perspective of health systems centers on cost containment, performance, and quality care. Digital health services routed in a data culture can help unlock productivity in health care. Platforms inspired by the sharing economy can bring tangible improvements in administrative automation, networked knowledge, and resource orchestration for higher productivity [12]. In 2015, Uber Health was able to deliver 2000 flu shots in 35 cities over 4 hours. However, countries in Europe still lag behind in patient empowerment and appreciation of network effects [13]. Collecting and analyzing data, health systems can measure the degree to which people are confident when navi-

⁴ www.estandards-project.eu

gating health services online, booking appointments or accessing their health information. Health systems may also share resources with other health systems to increase productivity.

Patient summary services linked to productivity and incentives can engage patients in their health goals in partnerships with the health team. Establishing incentives for high quality recording or assembly of patient summary data can directly improve data insights guiding interventions with direct impact on increased productivity, and patient satisfaction. Note that the concept of health system navigators is not new. It was first introduced by Harold Freeman in 1990⁵. Assisted by medical students, patients are able to navigate the logistical, emotional, and frequently cultural barriers of receiving care. Patients are assisting in assessing the situation and choices, articulate objectives, evaluate alternatives and reach decisions. In the end, healthcare is analog and human touch is paramount. Digital health tools can help fill the intention gap!

Perspective of health providers and the workforce

From the perspective of health providers and the workforce, maintain cutting edge competences is essential. The rapidly unfolding technological environment leaves health professionals perplexed and confused in front of *well-informed* or *arrogant know-all* patients that need to be “de-googled”. On the other hand, personalized treatment demands a partnership with the patient. For the workforce, the digital health compass needs to preserve the balance of roles in new quality relationships catalyzed by technology. Human relationships and empathy are challenged by complex knowledge-based remote cooperation. Self-care or care between care make physical encounters rare and precious.

Perspective of the digital health market

The digital health market is seeking opportunities to match demand and supply of digital health services. Established practices and a long standing *fee for service* reimbursement model of care, hinder adoption of innovation. Consider the case of the personal health record (PHR) as a digital organizer of personal health information. PHRs provide treatment support with self-management options. They facilitate sharing and exchange of health data with healthcare providers, suggest healthy lifestyle options, and track exercise, health and fitness. However, PHR adoption is not wide-spread and willingness to pay is low. Healthcare systems and government spending drive the market.

A recent study in 25 countries and 6 continents, revealed that most PHRs never go beyond the pilot stage of 100 -1000 subscribers. The study notes that some government strategies have succeeded in scaling up health information exchange between citizens and the health system. In contrast, the lifestyle focus of tech companies in the consumer market does not contribute to PHR initiatives across countries. Experience, preferences for engagement, health coaching, and use of digital health information are at an early stage. The promise of data fails to deliver market value. Exceptions to this rule marking future trends are patient advocacy groups that sponsor clinical trials and mediate prompt transfer of clinical results to routine patient care.

The eStandards project identified several gaps to be bridged by cooperative standards development including creating a reliable mix of patient and provider generated health data. Additional gaps identified were bridging health professional guidelines with clinical information models and terminology initiatives, establishing regulatory clarity, and attending to localization and adaptation of user requirements with a clear connection

to procurement. Lastly, eStandards highlighted the need for clear governance and maintenance of standards sets supported by open tools and data to leverage different speeds and cycles in sharing and integrated fragmented data. The bimodal practice of managing separate but coherent styles of work leveraging predictability and exploration is promising for diffusing health innovation matching technology supply with patient demand.

Perspectives of citizens and informal care givers

Citizens and households need to manage health information. Their methods vary in sophistication, time relevance, robustness, and creativity. Health information managers employ idiosyncratic tools and strategies to keep track of health data over time. Active involvement in health decisions is frequently expected and needed, as they navigate health systems to the best of their abilities seeking the best options. The first step to moving online is accessing and acting upon personal health information stored in health systems that are called to play a critical role in digital health literacy of both patients and the workforce.

Patients are best placed to have the most complete picture by being present in every encounter in a personal capacity. However, health confidence, health status, personal wellbeing, etc., combined with preferences for engagement, digital confidence, skills and capabilities including language, length, reading age, etc. affect their ability to use digital health tools and resources. Standards and quality labels reflect capabilities to connect and assess resources. Citizens make deliberate choices of sharing health related information, choosing to actively solicit, share or protect information pertinent to their conditions or health problem [14]. Patients want access to their health data, and appreciate opportunities to do so. Meanwhile, more and more, they share their experiences in online social networks and trust the feedback received.

With the Internet of Things, mHealth tools and apps add to the body of tools to manage data from the environment. IoT offers insights to the activities in the home, and the plethora of information sources and health related activities citizen engage in and perform [15]. Robust strategies to differentiate and handle health information emerge:

- *Just-in-time*, i.e. information and/or artifacts are with me at most times
- *Just-at-hand*, i.e. information and/or artifacts are visible or stored in readily accessible, highly familiar locations in a household
- *Just-in-case*, i.e., information and/or artifacts, either personal health files or general health information resources, are kept away, but are easily retrievable
- *Just-because*, i.e. information and/or artifacts of temporal relevance, kept in the household until storage strategy is assigned [16].

These human approaches to managing health information reflect strategies for maintaining confidentiality and privacy, bridge provider-generated and patient-generated data, and reduce fragmentation of health data providing support in the “care-between-care” period guiding inter-visit care actions [17].

Call for action: My data, my decision, our ePower

In this multifaceted health information ecosystem, our call for action is paramount. Adoption of innovations proceeds at the

⁵ <http://www.hpfreemanpni.org/>

speed of trust and issues of trust apply especially to the perspective of citizen and health providers. In our call for action we see collaboration and co-creation and where both digital health literate citizens and the health team see the relevance and value of accurate patient summaries. This implies that citizens become able and empowered to identify, appraise, and use their health information purposefully, and transform it into specific, trusted, and actionable knowledge applicable to the health reality at hand. Thus, health information should be relevant, regardless of source as health professionals or the citizen self are aligned to make sense of it in a sort of trialability that fosters large scale adoption, sharing and use [18].

My data

Individual knowledge and capability to manage one's health, to get appropriate help when needed and engage in shared decision-making are part of Health literacy. It emphasizes *motivation, knowledge and competencies to access, understand, appraise and apply health information* to all aspects of citizens' health judgments and decisions in everyday life [19]. Adding on the ability to appreciate and use productively digital health tools can be reflected in using patient summaries as the passport to this health journey.

My decision

Information relevant for the specific circumstances will help assemble information to participate in tomorrow's care. We expect more productive interactions and co-creation where the *personal dimension*; socio-demographic factors, Health and Digital Health Literacy span the full spectrum of "worried well" and the "really ill". *Digital literacy* boosts skills and capabilities to participate, and advance technology readiness. Marketing recognizes *Self Achievers, Priority Jugglers, Direction Takers, Balance Seekers*, and *Willful Endurers*. Taking into account the behavioral profile of an individual can help calibrate their compass. Moreover, the Health Confidence Score (HCS) is a short generic measure of a person's confidence to engage fully in their health and care measuring knowledge, self-management, access, and propensity to shared decision-making [20]. These are essential elements of advancing digital health literacy in a process that respects individuals allowing freedom of choice on when, how, and what digital health tools to use.

Our ePower

Empowering the citizens (patients, families and carers) to support a continuum of care across a range of services can relieve pressure on governments to provide more cost-effective and efficient health services improving health outcomes and encouraging citizens to manage their healthy life course. Beyond socio-demographic aspects that challenge adoption of digital health technology and a culture of sharing data that could drive routine evidence-based decision making.

Conclusion

The vision of the digital health compass as a navigation instrument that allows us to make sense of digital health innovations has been presented using patient summaries as an illustrative example. We argued that eHealth literate citizens are empowered and can make better choices. With increased digital health literacy, citizen can identify, appraise and use more health information resources, and transform them into trusted, actionable knowledge applicable to the situation at hand. Understanding one's health data is a foundation for empowerment. Co-creation and participatory design facilitate the design of information visualizations that are understandable and actionable, promoting active engagement. Cooperative use of standards is

the key to creating a trusted infrastructure for innovation, since innovation travels at the speed of trust.

Acknowledgements

This work has been funded in part by the European Commission by eStandards project under contract no: 643889. We also acknowledge discussions with and contributions from several colleagues who have helped shape these ideas and work directly or indirectly: Rita Medes, Henrique Martins, Morten Bruun-Rasmussen, Petter Hurlen, Petra Wilson, Céla Boyer, Ed Hammond, Suzanne Bakken, Tim Benson, Christian Lovis.

References

- [1] European Citizen's digital health literacy, Eurobarometer 404, 2014. http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf
- [2] J T Hart. The Inverse Care Law, *The Lancet*, **297** (1971) 405–412
- [3] CD Norman, Skinner HA, eHEALS: The eHealth Literacy Scale. *J Med Internet Res*. **8** (2006) e27.
- [4] K Jethwani, J Kvedar, J Kvedar. Behavioral Phenotyping: A Tool for Personalized Medicine. *Personalized Medicine*. **6** (2010) 689–93
- [5] D deBronkart, E J Topol, (2013) *Let Patients Help!*. <http://www.epatient-dave.com/let-patients-help/>
- [6] EU Commission. (2014). Green Paper on Mobile Health ("mHealth"). SWD (2014) 135 final. EU Commission.
- [7] M Hostetter, S Klein, D McCarthy. (2014). *Taking Digital Health to the Next Level: Promoting Technologies That Empower Consumers and Drive Health System Transformation*. The Commonwealth Fund.
- [8] S de Lusignan, F Mold, A Sheikh, A Majeed, J C Wyatt, T Quinn, M Cavill, T Gronlund, C Franco, U Chauhan, H Blakey, N Kataria, F Barker, B Ellis, P Koczan, T N Arvanitis, M McCarthy, S Jones and I Rafi. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. *BMJ Open* **4** (2014) e006021
- [9] M Hostetter, S Klein, D McCarthy. (2014). *Taking Digital Health to the Next Level: Promoting Technologies That Empower Consumers and Drive Health System Transformation*. The Commonwealth Fund.
- [10] O Heinze, M Birkle, I Köster and B Bergh. Architecture of a consent management suite and integration into IHE-based regional health information networks. *BMC Medical Informatics and Decision Making* **11** (2011) 58
- [11] C Boyer, M Selby, J R Scherrer, R D Appel. The Health On the Net Code of Conduct for medical and health Websites. *Computers in Biology and Medicine*. **28** (1998) 603–10
- [12] J Bush, J Fox. Bring the Power of Platforms to Health Care. HBR November 2016
- [13] J Zelmer, E Ronchi, H Hyppönen, F Lupiáñez-Villanueva, C Codagnone, C Nohr, U Huebner, A Fazzalari and J Adler-Milstein. International health IT benchmarking: learning from cross-country comparisons. *J Am Med Inform Assoc*. **24** (2017) 371–379
- [14] M van der Velden and K Emam,. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. *J Am Med Inform Assoc*, **20** (2012) 16–24.
- [15] T Zayas-Cabán. Health information management in the home: A human factors assessment. *Work* **41** (2012) 315–328.
- [16] A Moen, P F Brennan. Health@Home: the work of Health Information Management in the Household (HIMH) - implications for Consumer Health Informatics (CHI) innovations. *J Am Med Inform Assoc*, **12** (2005) 648–56.
- [17] P F Brennan, G Casper. Observing health in everyday living: ODLs and the care-between-the-care. *Personal and Ubiquitous Computing*, **19** (2014) 3–8,
- [18] E M Rogers. *Diffusion of Innovations*. The Free Press, New York 1962/1995.
- [19] K Sørensen, VS an den Broucke, J Fullam, G Doyle, J Pelikan, Z Slonska and H Brand . Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health* **12** (2012) 80
- [20] T Benson, C Bowman and H W W Potts. Health Confidence Score (HCS): development and validation of a short generic questionnaire for person-centred care. (under review)

Address for correspondence

Catherine Chronaki, 38-40 Square de Meeus, Brussels, 1000, Belgium.
eMail: euoffice@HL7.org

Electronic Health Records in the Cloud: Improving Primary Health Care Delivery in South Africa

Liezel Cilliers^a, Graham Wright^b

^a Department of Information Systems, University of Fort Hare, East London, Eastern Cape, South Africa,

^b Department of Health Sciences, University of Fort Hare, East London, Eastern Cape, South Africa

Abstract

In South Africa, the recording of health data is done manually in a paper-based file, while attempts to digitize healthcare records have had limited success. In many countries, Electronic Health Records (EHRs) has developed in silos, with little or no integration between different operational systems. Literature has provided evidence that the cloud can be used to 'leapfrog' some of these implementation issues, but the adoption of this technology in the public health care sector has been very limited. This paper aims to identify the major reasons why the cloud has not been used to implement EHRs for the South African public health care system, and to provide recommendations of how to overcome these challenges. From the literature, it is clear that there are technology, environmental and organisational challenges affecting the implementation of EHRs in the cloud. Four recommendations are provided that can be used by the National Department of Health to implement EHRs making use of the cloud.

Keywords:

Electronic health records; cloud computing; primary health care

Introduction

The primary health care system in South Africa is facing many challenges. These difficulties include the high disease burden consisting of infectious and non-communicable diseases, persisting social disparities, and a lack of qualified health staff. South Africa's population is currently estimated in approximately 54 million people, while 54% of this population lives below the poverty line [1]. It has been reported that in South Africa 84% of the population cannot afford a medical aid, and thus making use of the public health care sector, which only employs 20% of the health care workers in the country [2].

The shortage of primary health care facilities is most acute in the rural areas of the country. Patients living in these areas often travel far distances at considerable cost to access health care services [3]. When the primary care facility cannot treat the patient, they are referred to the next tier in the health care system, which is the secondary level district hospital, or the last tier, that provides tertiary services, and these are often located far away in the city. Thus, patients can have many health care providers throughout the public health care system, all of which will have their own health care records for the particular patient. If the file is misplaced at the health care facility, a new file must be opened and all the medical history before the event is lost. This means that patients often have multiple files at a single facility. To exacerbate the situation, if the patient move to another region or province, their health record will remain behind, and their health care record will have to be started anew [4].

The most common method to record health care data in South Africa is a manual, paper-based health records. The National Department of Health (NDoH) has made attempts in the past to implement an electronic health record system to improve the continuity of care, but these efforts were fragmented at best resulting in interoperability issues between the various systems being used [4]. While the cloud can overcome the problem of interoperability as one system can be made available to every health care facility in South Africa, the adoption of the cloud in the public health care sector in the country has been slow. Some of the reasons provided for the slow adoption of the cloud are security concerns, trust of the cloud, and compliance to legislation [5]. The aim of this paper is two-fold: First, the challenges why the cloud has not been used to implement EHRs in South Africa will be investigated, and second, recommendations to overcome these obstacles will be provided. The next section will discuss the definition and use of electronic health records in the South African public health care system.

Electronic Health Records

According to the Health Information and Management System Society [6], although the terms EHRs and electronic medical record (EMR) are often used interchangeably, they actually refer to separate concepts. An EMR is a local, stand-alone health information system that allows for the storage, retrieval and modification of health records. The EMR refers to the legal patient record that is created in the health care environment, by the health care worker, and becomes the data source for the EHR.

The EHR is an electronic version of the patient's medical history, which includes clinical data and history, progress notes, medications, vital signs, past medical history, immunisations, laboratory data, radiology reports, and administrative data. The EHR can be shared among health care providers at the same hospital and across different health care settings through an information system that is connected via a network, e.g. the cloud [3, 5]. However, many patients have privacy concerns when their medical data is stored in an EHR.

Privacy in the context of EHRs means that the patient must retain control of their own information even if it is 'owned' by another party, such as the hospital. This control extends to whom is authorized to access a copy of the data, e.g. where billing records are shared with the medical aid company, or different health care workers are authorized to view the EHR [7]. The EHR must be delivered in an environment in which patients have absolute confidence that their privacy will be protected and that the confidentiality, integrity and availability of the information stored in EHR will be assured [3]. The next section will discuss cloud computing in the healthcare sector.

Cloud Computing

Cloud computing has the potential to make health care more efficient and effective. Storage of medical data, processing capabilities of the EHRs, and bandwidth necessary to share the records are centralised to become more efficient. However, while the cloud has been adopted in many industries, the health care sector has been slow to make use of cloud-based solutions [5, 8].

The cloud makes use of internet connectivity and a standard web browser to allow the health care worker to access the EHR. An advantage of the cloud is that the health care organisation do not have to develop or maintain the hardware and software infrastructure necessary for the EHRs, but simply make use of the cloud services and deployment models that best suit their specific operational and technical needs [5].

There are three service models that are applicable to the cloud in the health care sector. The first is Infrastructure as a Service (IaaS) that allows health care workers to provision processing, storage, network, and other resources in which they can deploy and run arbitrary software, such as operating systems and EHRs applications. This means that the health care organisation does not have to manage their own infrastructure and can add computing processing power and storage capacity as the need arise. The main benefit of IaaS is the fast deployment of applications and the improved agility of information technology services within the organisation. However, the NDoH do not have their own EHRs application in place at present. Therefore, this option does not seem appropriate. The second service model is the Platform as a Service (PaaS) that allows for the quick and efficient development of health applications that can be hosted in the cloud. This environment allows the NDoH to build applications with no capital expenditure on infrastructure, as the service provider is responsible for the maintenance and control of the underlying cloud infrastructure. The NDoH could use this model to build their own EHRs, but with the lack of information technology skills in the public health care sector in South Africa, this solution does not seem viable. Third, service providers make use of a Software as a Service (SaaS) model to provide access to applications and software remotely through a web-based service. The cost of accessing these services is typically less than paying for licensed applications, as the health care organisation only pay for the actual time they use the application, and there are no hardware layout costs. The service provider is also responsible for the installation, set-up and maintenance of the services they provide. In this model, the service provider would provide the application for the EHRs, and the NDoH would pay for the usage thereof [9; 10]. This option would also overcome the problem of integration between the various health care facilities, as a single program will serve the South African public health care system.

There are four cloud deployment models: private cloud, community cloud, public cloud, and hybrid cloud. The private cloud is hosted and operated by one organisation, while the community cloud allows several organisations, normally with a common goal to share the infrastructure. The public cloud is owned by an organisation that sells cloud services to the general public or a large industry group. The last deployment model, the hybrid cloud, is a composition of two or more clouds that remain unique entities but are bound together by proprietary technology [10]. The private cloud could be used by one organisation, such as a medical aid company, to provide EHRs to their clients, but this system would not be integrable with other systems outside the company's sphere. A company such as Google that provides EHR services to the public would make use of a public cloud, but there are privacy and security concerns with such a deployment model. The model that is best

suited to EHRs, therefore, seem to be a community cloud, as health care providers share a common goal – to provide more efficient and effective access to the medical data of the patient [9, 10].

The benefits of cloud computing to provide EHRs to the public health care service in South Africa can be derived from the discussion above as well as the following [11]:

1. Reduced cost to the NDoH, as they do not have to maintain the information technology infrastructure.
2. The cloud is available on demand to the health care worker and can be accessed from any location.
3. The resource pooling of physical and virtual resources allows for the most efficient use of the technology.
4. The cloud provides flexibility to the NDoH, as it can be easily scaled according to the need and demand of the health care facility.
5. The cloud provides usage metrics that can help the NDoH in the decision-making process for future technology usage and planning. These metrics can also contribute to the sustainability of the EHRs that is provided.

Electronic Health Records in South Africa

HealthID

Discovery Medical Aid is providing an EHR to their clients called Health ID. This EHR is the first of its kind in South Africa according to Discovery. The purpose of HealthID is to collate the medical data and details of all hospital visits, including scripts for medicines, blood tests and health indicators. The patient must provide written consent for each of their health care workers they visit to access their HealthID in order to protect their privacy. The consent can be provided before the appointment or during the visit by accessing the patient's Health ID profile on the Discovery website [12].

The purpose of Health ID is to reduce the administrative burden on the health care worker, as they have access to the full medical history of the patient regardless of where their previous visits took place, as well as the health benefits of the patient. The real-time access allows the health care worker to apply and receive approval for chronic medication and to identify what medication the patient is currently taking, which is especially useful for elderly patients that may not be able to remember [12].

Google Health

This product was available in South Africa through the Google platform from 2008-2012. Google Health provided the patient with an opportunity to manage their own health information, although the EHR could not be shared with the health care worker. Unfortunately, Google Health was not adopted by the general public and achieved only limited use; hence, it was discontinued in 2012. While the system quality of Google Health was found to be appropriate, the system was perceived to be high risk resulting in low trust among users. The lack of awareness among users about the benefits of EHRs contributed to the lack of trust, as well as the failure of health care workers to recommend the product as a useful tool to collect health information. The functionality of Google Health was also limited. The user could not make the application available to their health care provider, but only use it to manage their own health information. Klein suggested that patients were not ready to manage their own health data in 2008, which contributed to the lack of adoption among users [13].

Nompilo service

Vodacom in partnership with local Non-Governmental Organizations and the NDoH launched an application which provided end-to-end monitoring and evaluation for local

community workers in 2009. When these workers visit the patients in the community, they can upload patient information directly to the system making use of their smartphones. Each patient is assigned a unique 2D barcode that can be used to identify them on the system. Nompilo then shows the health care worker what information needs to be collected from that particular patient. Typical information collected include personal, administrative, and physiological data to support the treatment plan. Current and historical patient information can also be viewed remotely, in real time, by the health care worker [14].

Electronic health record systems in the public health care sector

The South African NDoH has also started to implement EHR systems in their public health care sector. Currently, 5 out of the 9 provinces in South Africa have some form of EHR system operational in public hospitals. In the KwaZulu-Natal province, some hospitals use the Medicom or Meditech EHR system, while in the Western Cape province a few hospitals use the Unicare or Clinicom EHR systems. Hospitals in the Limpopo province also use the Unicare or Medicom EHR systems [15]. The implementation of different EHR systems from various vendors presents a challenge, as these systems are built with different underlying database architectures; and therefore, often fail to communicate and share information amongst each other. Also, while these systems have been implemented in some areas, the majority of the public health centres in South Africa still make use of a paper-based filing system [16].

Methods

The research was exploratory in nature, and reviewed literature from previous studies, e.g. articles published in academic journals, books (both print and electronic formats), conference proceedings and websites relative to this study. Electronic databases such as ACM Digital Library, Sage Online Journals, Science Direct, Springer Link and Sabinet Reference were used to find the literature. Research keywords such as 'electronic health record', 'electronic medical record', 'cloud' and 'primary health care' were used as search terms. The reference lists within articles previously identified were also examined to yield additional literature. Inductive reasoning was applied to analyze the literature making use of a thematic content approach. The results of the analysis are discussed in the next section.

Discussion

Duhaime [17] provided the following list of generic barriers that must be considered when cloud computing is implemented in any organisation (Figure 1). These barriers are discussed in the context of EHRs below.



Figure 1 - Generic barriers to the implementation of cloud computing (adapted from Mathew, 2012 [18])

Scale, Cost and Reliability

Performance concerns about the cloud are often based on the scale, availability and the dispersal of the cloud [18]. The technology needs to be reliable and accessible to be useful to the health care worker. In Africa, the primary barriers to the implementation of the cloud include unreliable electricity supply, poor internet connectivity, and limited bandwidth. Internet penetration in Africa is the lowest of all the developing world regions, with only 0.3% fixed broadband penetration on the continent, and internet penetration at 16.3% [19]. As an internet connection is needed to access the information in the cloud, these factors will influence the reliability of the EHRs. There are several different factors that will determine the cost of the technology, especially if SaaS service model is used. These include the amount of storage required, network services, scheduling, service level agreements, optimal location of data centres and software components, efficient SQL query processing, architecture and process improvement.

Lock In and Agility

One of the advantages of the cloud is the agility that it provides in regards to capacity, storage and demand. When the NDoH decides to make use of one vendor, they become 'locked-in', as they are then obliged to deal only with that company and their products. If there is a need to migrate away from that vendor, then a significant cost will be incurred. At the present moment, the EHR systems that are used within and across provinces are not integratable [16]. The patient data in these legacy systems will have to be imported into the new EHRs to ensure continuity of care. This may be a problem as the maturity of the technology in South Africa is still low, which means that the industry is not regulated and standardisation between products are non-existent [3].

Change Management

Cloud computing is a new technology that is changing the traditional way of managing health care information in the public health sector. The lack of knowledge about the cloud is one of the reasons the NDoH is reluctant to migrate to the cloud because health care staff and patients are not aware of the cloud's purpose [20]. Due to this lack of understanding, staff may fear that they will become redundant and patients may have misgivings about the safety and privacy of their information and resist the use of the new technology [4].

The failed Google Health application was a prime example of how the relationship between knowledge of the cloud and trust in the cloud is essential for the adoption of the technology. This is because knowledge of the technology is likely to enhance the levels of trust of the user, ensuring it's continued use (Google). Therefore, the NDoH must conduct a needs analysis of their employees' and patients' knowledge of cloud computing before the technology is implemented to manage patient records [10].

Lack of Control

EHRs have to manage sensitive patient information. Comminos [21] found that one of the main reasons why patients and healthcare workers are reluctant to make use of the cloud is concerns about third-party cloud providers that must store or process this sensitive information. Also, connectivity issues are also a concern, as the health care worker can only access information if they are connected to the internet. The lack of

transparency in the cloud is contributing to the control concerns of EHRs. The user is expected to upload their health care information without any knowledge regarding the security measures or controls in place in the cloud. Often, the user does not even know where the information is located, making litigation difficult if something should go wrong [20]

Security

EHRs store and process large databases of sensitive patient information. Many patients have expressed privacy and security concerns if this data is transferred to third party providers of cloud services as the cloud is seen as inherently not secure [20]. The main reason for this perception is that the security in the cloud is intangible and less visible, which creates a concern whether information is stored securely. Security concerns include the altering or loss of sensitive data in the cloud and the unauthorised use of the data by cloud providers [18].

Discussion

In the previous section, the generic barriers for the utilization of the cloud to provide sustainable EHRs services in the public health care service of South Africa were discussed. In summary, the following critical areas were identified:

- Technology: Vendor lock-in; agility and maturity of products available; standardization, reliability and availability of technology; cost of making use of the cloud;
- Organization: Change management including trust; adoption and awareness of the cloud; and
- Environment: Security and privacy concerns.

The following are recommendations that were developed to overcome these critical areas and improve the implementation of EHRs in the cloud. The first recommendation is to provide education to improve the level of awareness about and trust in the cloud, and the benefits it can provide to EHRs. One of the reasons why Google Health failed was because it did not engage with the users showing why EHRs was beneficial to manage their health care, and how cloud computing could help them to manage their EHR. The recommendation is thus twofold as both health care workers and patients need to be educated about the advantages and potential risks of making use of the cloud. With this knowledge gain, it should increase user acceptance and utilization of the technology.

The second recommendation is for the government to provide best practices for both EHRs and cloud computing within a legislative framework. EHRs are already included in the E-Health Strategy for South Africa, but the National Department of Health must also provide a national integrated plan for the implementation of a national EHR system. Best practices, such as to how any service can be implemented making use of cloud computing will also address the problem of vendor lock-in and legacy systems that are not integrable with the national system. Legislation and policies need to be put in place to ensure compliance of the technology with international standards.

The third recommendation is to improve the transparency of the cloud to address the privacy and security concerns. Legislation that provides a proper legal framework, such as the Protection of Personal Information Act, will guide the user, health care worker and cloud provider as to what their rights and responsibilities are. This improves the control on the part of the user and health care workers.

Security controls in place, to protect data, should be provided by third-party cloud providers to assure users that their data is

securely stored. Security controls will further improve the transparency of the cloud and provide peace of mind to the patient that the privacy of their sensitive information will be protected. The recommended security requirements include certification, where developers open their cloud applications to security specialists that provide certification for security. Certification will certainly improve the trust of the user in the cloud-based EHRs [22].

- Identification and Authentication: all accounts should incorporate usernames and passwords to verify the user on the cloud, and provide an audit trail of access to that specific EHR;
- Authorization: The challenge is to provide the right information to the right person, while limiting the access privileges of the user. Patients should be able to grant privileges to health care workers, while the functionality of the EHR should be restrictive to who can add or modify information. Patients may not realize why it is important not to change test or laboratory scores. Each user will have to be allocated manually with security clearance to ensure limitation of access to different files;
- Integrity and Confidentiality: encryption techniques should be used to protect sensitive data of the patient;
- Non-repudiation: health care workers and patients must have access to the data at all times for EHRs to be useful. Techniques including digital signatures, timestamps and confirmation receipt services can be deployed to achieve non-repudiation; and
- Availability: It means that the data will be available, accessible and obtainable at all times on the health cloud, while time-outs must be used to restrict access of unauthorized persons if the user forgets to log out of the application.

The fourth recommendation is to address technology problems that may affect the cloud specifically. Cloud computing is still a relatively new technology with limited use in the health sector. Specific technology factors that must be considered include the reliability, cost and availability of the technology. These are important because access to health data must be reliable, if the EHRs is to be of any use. The total cost of migrating to the cloud will decrease, as more deployment options become available. The cost of bandwidth, which is very expensive and limited in South Africa, must also be decreased to make this a viable technology. In developing countries, especially where IT infrastructure is poorly developed, cloud computing makes use of existing internet connections without requiring more infrastructure or financial investments.

Conclusion

This paper provides an overview of the current use of EHRs in South Africa. This includes systems available to the individual patient and the health care worker to manage medical data, and illustrates that EHRs are slowly making inroads in the health care sector. Reasons why the cloud has not been used to provide EHRs in the public health sector of South Africa were divided into the organizational, environmental and technology categories. Four recommendations were provided to how to overcome these challenges that can be used by the NDoH to implement cloud-based EHRs.

Future work would include testing these recommendations making use of an expert review to refine them. It would be useful to investigate the extent of EHRs implementation in the public health care sector and compare these results with the private health care sector. An examination of the adoption of

HealthID among patients belonging to the Discovery medical aid could also provide valuable insight.

Acknowledgements

This research project was jointly funded by the South African Medical Research Council (SAMRC) and Forte, the Swedish Research Council for Welfare, Working Life and Welfare.

References

- [1] Statistics South Africa, Mid-year Population estimates (2014). Available at: www.statssa.gov.za/publications/.../P03022014.pdf (accessed 1 November 2016).
- [2] BM. Mayosi and SR. Benatar, Health and health care in South Africa--20 years after Mandela, *The New England Journal of Medicine* **371**(14) (2014), 1344–53.
- [3] M. Katurura, and L. Cilliers, The extent to which the POPI Act makes provision for patient privacy in mobile personal health record systems, In the proceedings of *IST-Africa*, Durban, 3-5 May 2016, South Africa.
- [4] D. O'Mahoney, G. Wright, and L. Cilliers, Electronic health information systems for primary care in South Africa: a review of current operational systems. In the proceedings of *Information Communication Technology and Society Conference*, Umhlanga, 1-3 May 2017, South Africa.
- [5] J. Puustjarvi, and L. Puustjarvi, The Opportunities of Exploiting Cloud Computing in Telemedicine: A Developing Country Perspective, In the proceedings of *IST-Africa*, Tanzania, 1-3 May 2012, Tanzania.
- [6] HIMSS, Available at: <http://www.himss.org/ASP/index.asp>. (accessed 10 November 2016).
- [7] S. Avancha, A. Baxi and D. Kotz, Privacy in mobile technology for personal healthcare, *ACM Computing Surveys (CSUR)* **45**(1) (2012), 3.
- [8] D. Chappel, A Short Introduction to Cloud Computing: An Enterprise-Oriented View. Available at: <http://digitalebookden.com/a-short-introduction-to-cloud-platforms.html> (accessed 10 November 2016)
- [9] S. Subashini, and V. Kavitha, A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications* **34**(1) (2012), 1-11.
- [10] B. Baize, In cloud we trust. *Cloud computing* **16** (2011), 21-30.
- [11] G. Murjithi, and J. Kotze, Cloud computing in higher education: implications for South African public universities and FET colleges. *14th annual conference on World Wide Web applications*, Durban, 2012, 4-24.
- [12] Discovery, Discovery Health launches HealthID, 2012.
- [13] T. Spil, and R. Klein, Personal Health Records Success; Why Google Health failed and what does that mean for Microsoft HealthVault? In the proceeding of the *47th Hawaii International Conference on System Science*, Hawaii 2012.
- [14] Mezzanine, Vodacom works with the SA Ministry of Health and local NGOs to provide better healthcare, 2014.
- [15] J.E. Ataguba, and D. McIntyre, Paying for and receiving benefits from health services in South Africa: Is the health system equitable? *Health Policy and Planning* **27**(SUPPL.1) (2012), 35–45.
- [16] D. O'Mahony, G. Wright, P. Yogeswaran, and F. Govere, Knowledge and attitudes of nurses in community health centres about electronic medical records. *Curationis* **37**(1) (2015) Art.#1150, 6 pages.
- [17] D. Duhaime, The Pros and Cons of Cloud Computing. In the Spotlight, 2011.
- [18] S. Mathew, Implementation of cloud computing in education - a revolution. *International Journal of Computer Theory and Engineering* **4**(3) (2012), 473-476.
- [19] M. Mars, Telemedicine and advances in urban and rural healthcare delivery in Africa. *Progress in cardiovascular diseases* **56**(3) (2013), 326-335.
- [20] F. Koch, M. Assuncao, and M. Netto, *A cost analysis of cloud computing for education*. Berlin: Springer, 2012.
- [21] A. Comminos, Southern African Internet Governance Forum Issue Papers. Johannesburg: SANGONeT for the Southern African Internet Governance Forum, 2011.
- [22] S. Ramgovind, and M. Eloff, The management of security in cloud computing. *Information Security for South Africa (ISSA)*, Johannesburg: ISSA, 2010, 1-10.

Address for correspondence

Corresponding author: Liezel Cilliers.
Email: lcilliers@ufh.ac.za or
Phone: +027 43 7047067.

Decision-Tree Model for Support of Health Policy Choices Based on Pneumococcal Conjugate Vaccine (PCV) Program Outcomes

Andre Cohen^a, Casimir A. Kulikowski^a, Timothy Elbert^a,
Samantha Clark^c, Dagna Constenla^c, Anushua Sinha^b

^aDepartment of Computer Science, Rutgers-The State University of New Jersey, Piscataway, New Jersey, USA

^bMerck Research Laboratories, Rahway, New Jersey, USA

^cDepartment of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Abstract

Pneumococcal Conjugate Vaccine (PCV) has the potential to save lives in low-income countries. We have developed a computational model and web-based decision support software for comparing cost-benefit tradeoffs from alternative PCV program designs, considering their direct and indirect effects on early childhood populations in resource-poor settings. This supports policy-makers in estimating potential health outcomes and cost-effectiveness of different vaccination program strategies for a wide range of population coverage and vaccine effectiveness assumptions.

Keywords:

Health Policy; Conjugate Vaccines; Decision Support

Introduction

Despite its potential to significantly reduce pneumococcal disease – a major cause of death in under-five year olds in resource-poor regions of the world, the high price of Pneumococcal Conjugate Vaccine (PCV) has deterred introduction in many of the low- and middle-income countries where such prevention is needed the most. For policy makers in these countries, often operating in resource-constrained environments, the relative costs and benefits associated with introduction of a vaccination program as a one-time effort vs. one with catch-up for children who may have missed the first opportunity and may require catch-up vaccination, is something that needs to be estimated rapidly and efficiently. Traditional epidemiological models require detailed estimation and computation – a review of different modeling approaches can be found in the literature [1]. For practical purposes of providing decision-support for making policy recommendations, an incremental cost-effectiveness model, such as that developed in the TRIVAC system [2] as a spreadsheet software program, has been shown to be most effective to obtain estimates for policy decisions. It uses parameters, such as demography, disease burden, vaccine costs, vaccine coverage, vaccine efficacy, health service utilization, and costs. With results of this kind of model in mind, we have developed an interactive, portable web-based health policy analysis software tool [3] that is computationally efficient, and allows calculations and comparisons with alternative assumptions about population characteristics and vaccination program details from a wide range of investigations [4 – 11] about the effects of pneumonia vaccination, which we have designed for estimating impact of the introduction of vaccination programs to counter infant pneumonia in countries with limited resources.

Methods

Our model is based on combined results from three international efforts to assess the burden of pneumococcal disease and to develop standards for assessing cost-effectiveness of preventive vaccination strategies: a) WHO-Hib Pneumococcal Global Disease Burden Project (GDB) for children under 5 [2]; b) Global Serotype Project (GSP) for regional proportions of disease-causing pneumococcal isolates covered by PCVs; and c) WHO CHOICE for country-level medical costs of disease management and treatments.

Decision Tree Model Structure and Parameters

The model divides a birth cohort into subgroups of patients who experience pneumococcal pneumonia, pneumococcal meningitis, and non-pneumonia/non-meningitis invasive pneumococcal disease (IPD), with each subgroup distributed over possible outcomes of death, disability (deafness, seizures, etc.) or survival without disability. A decision tree structure is

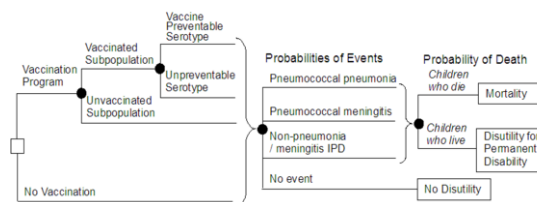


Figure 1 - Decision Tree Model for Vaccination Decision

used with contingent events appearing as branching sub-trees, and possible outcome states (terminal nodes) associated with utilities, or Disability Adjusted Life Years (DALYs).

Branches of the tree, following round nodes, may occur with given probabilities which are used in this model to represent vaccination rate, serotype coverage, or the chances of mortality and morbidities, etc. Each branch of the tree may also be associated with costs, which accumulate downstream. The basic pneumococcal vaccination policy model, illustrated in Figure 1, involves a decision between two scenarios: Vaccination Program and No Vaccination. In the vaccination scenario, a given proportion of the population is vaccinated, which in turn can prevent a proportion of the serotypes of disease. Probabilities of events vary for each subgroup of the population in each scenario, which can be computed based on the prevalence of each event and the vaccine efficacy, if any, for those events. Children with disease have a certain chance

of death, and those who survive may have an amount of disability over the course of their remaining life expectancy.

Disease Incidence and Case Fatality Ratios

The probabilities of events are based on data available from the GDB project, including disease incidence (per 100,000 infants) for each country, estimates for case fatality rates, the number of events, and deaths observed in children under 5, that were subject to a rigorous meta-analysis. GDB incidence reflecting a 5-year cohort is converted to a probability of disease in a single birth cohort.

Case fatality ratios are also derived from the GDB data; however, since the incidence values for non-pneumonia/non-meningitis IPD reported in the GDB are based on acute disease, we had to explicitly re-compute the incidence based on the underlying values for all disease. For this reason, we also accept the observed number of events and deaths as inputs, as an alternative to the case fatality ratio, which is provided if available.

Vaccine Serotype Coverage and Efficacy

In the vaccination scenario, the probabilities of events for vaccinated children reflect the impact of vaccine efficacy for the covered serotypes. This is further complicated by modelling the varying efficacy in presence or absence of HIV, so resulting efficacy can be expressed by the formula:

$$\text{Effective Efficacy} = \text{Vaccination Coverage} * \text{Serotype Coverage} * [(\text{Prevalence of HIV}) * (\text{Efficacy with HIV}) + (1 - \text{Prevalence of HIV}) * (\text{Efficacy without HIV})]$$

We used 2006 WHO/ UNICEF estimates of the proportion of children receiving three doses of diphtheria-pertussis-tetanus vaccine (DTP3) to estimate the proportion of children likely to receive three doses of pneumococcal conjugate vaccine when introduced into a childhood immunization schedule. Serotype coverage estimates were initially estimated as the proportion of disease-causing pneumococcal isolates that are serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F in children under-five. These serotype coverage estimates were taken from regional estimates from the Global Serotype Project [GSP Summary Report 2007]. The regional groupings used were Africa, Asia, Europe, Latin America, the Caribbean, North America, and Oceania. The online edition of the model was updated to use estimates for a 10-valent vaccine (1, 4, 5, 6B, 7F, 9V, 14, 18C, 19F, 23F) computed from the meta-estimates that appear in GAVI's Pneumococcal ADIP TPP Codebook.

Measure of Effectiveness: Disability adjusted life years (DALY)

Disability Adjusted Life Year or DALY is a burden of disease measure used to assess the relative utility of each possible health outcome. The DALY combines in one measure the time lived with disability and the time lost due to premature mortality. One DALY can be thought of as one lost year of 'healthy' life. It is calculated based on the age at which disease occurs, the life expectancy at the time of disease occurs, and the expected duration of disability. It is weighted to account for the years of life during which an individual is anticipated to be most economically productive and is discounted to present value. Years of Life Lost (YLLs) is computed as a function of a discount rate (r), an age-weighting parameter, an age-weighting modulation factor, a constant, and the life expectancy at the age of onset of disability. Standard values from the WHO GBD project are used as defaults, but can be adjusted by the user. When calculating disability-adjusted life years

attributable to acute illness or non-fatal permanent sequelae, it is assumed that time lived with acute illness or with a sequelae is valued less than time lived in perfect health. Disability weights are used to represent the preference for time lived in perfect health relative to time lived with acute illness or with a permanent sequelae. DALYs attributable to acute illness or permanent sequelae are termed Years Lived with Disability (YLDs) and are calculated using the formula:

$$YLDs[r, K, \beta] = \text{Disability Weight} * YLLs[r, K, \beta]$$

where the disability weights for each sequelae are also the standard values from the WHO GBD project.

The summary metric the model aims to calculate is the incremental cost-effectiveness ratio. This is:

$$\text{Net costs/ Net health benefits} = (\text{Vaccine-related costs} - \text{Averted disease costs}) / (\text{Disability-adjusted life years averted})$$

Data from diverse sources are drawn together in calculating each of the three terms. The quality of the completed model is dependent on the quality of the data input into the model. The development team has striven to identify the highest quality sources of international data for use in the pre-populated model. If modifying these pre-populated inputs, we encourage users to develop a quality assurance process, to consider accuracy, precision, and generalizability of estimates prior to incorporation into the model.

The Web-Based Interface and Software Design

The model is embedded in a general purpose decision support framework consisting of a series of dynamic web pages that provide complete documentation of all input parameters to the model, and a mechanism to allow policy makers to enter new values within reasonable limits. It is operable on any standards-compliant web browser. For convenience, pre-populated values are available for each country, for which GAVI eligible countries can be selected on a map. The input data values for the model are presented in series of pages, arranged as a "tabbed" notebook, in which they can be examined or modified. Most entries, technical term, and sources of baseline data values, are documented by hyperlinks to Glossary, Frequently Asked Questions, and References pages. An example interface page is shown in Figure 2.

Costs are in Afghani		Baseline	Sources
Hospital Costs			
Cost per bed day at a primary level hospital	632.37	632.37	WHO CHOICE 2009
Cost per bed day at a secondary level hospital	894.91	894.91	WHO CHOICE 2009
Cost per bed day at a tertiary level hospital	144.42	144.42	WHO CHOICE 2009
Yearly cost discount rate	0.03	0.03	WHO GBD Project (Barry 1998)
Direct Medical Hospital Costs			
per case			
Outpatient Costs			
Cost per outpatient visit at a primary level hospital	147.81	147.81	WHO CHOICE 2009
Cost per outpatient visit at a secondary level hospital	209.7	209.7	WHO CHOICE 2009
Cost per outpatient visit at a tertiary level hospital	319.36	319.36	WHO CHOICE 2009
Cost per 100 consults at a secondary health center	207.9	207.9	WHO CHOICE 2009
Direct Medical Outpatient Costs			
per case			
Direct Non-Medical / Social Costs			
Optional []			
Average minimum daily wage rate for family caregivers			
Number of days of work lost for family caregivers for pneumonia	2.5	2.5	(Average hospital stay)
Number of days of work lost for family caregivers for meningitis	2.5	2.5	(Average hospital stay)
Number of days of work lost for family caregivers for non-pneumococcal non-meningitis IPD	2.5	2.5	(Average hospital stay)
Number of days of work lost for family caregivers for all cause acute media	2.5	2.5	(Average hospital stay)
Family transportation costs for pneumonia	0	0	
Family transportation costs for meningitis	0	0	
Family transportation costs for IPD	0	0	
Family transportation costs for acute media	0	0	
Family out-of-pocket medication and diagnostic costs of pneumonia	0	0	

Figure 2 - Example of web-page with data entry choices

Table 1 - Cumulative Disease Burden for Vietnam

	No vaccine		Vaccine		Under 2 Catchup	
	Cases	Deaths	Cases	Deaths	Cases	Deaths
Pcc Pn	1,312,245	7,873	273,519	1,641	230,359	1,382
Pcc Mn	4,971	1,437	919	265	818	236
NP-NM IPD	29,173	1,433	6,080	298	5,121	251
Total S.P.	1,346,390	10,743	280,520	2,205	236,298	1,870

Software features for efficiency and portability include:

- Use of standard HTML pages, making the decision support framework platform-independent and operable on any computer with a web browser,
- The interface and evaluable model can be accessed in areas without internet access via CD-ROM,
- Ability of users to perform customized sensitivity analysis,
- Can import easy-to-use Excel spreadsheets that contain all the variables and values for specific countries into the model, and
- Provides an efficient, low-cost method for geographically dispersed analysts to project and compare disease burden and cost outcomes for vaccination strategies of interest for specific countries or regions utilizing the best available data.

Results

Key model outputs can be prepared by country, by year, or cumulative over ten years, and include:

- Pneumonia, meningitis, and NPNM IPD Cases/deaths and cases/deaths averted,
- YLDs, YLLs, total DALYs, and DALYs averted,
- Vaccine program costs, disease treatment costs and net costs, and
- Average cost-effectiveness of a vaccination program (each option compared to no vaccine) and incremental cost-effectiveness (options ranked and compared to next best option) over the years of a program.

A screenshot of results from running the model for the country of Vietnam are shown below in Figure 3, highlighting the cumulative disease burdens after 10 years of a vaccination program, and the reduction in cases of pneumonia and expected deaths contrasting introducing a newborn vaccination program only vs. alternatives of programs with catch-up from enrolling children under 1 of age or under 2 years of age.

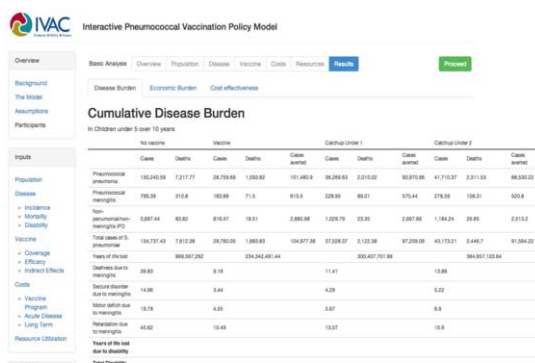


Figure 3 - Cumulative Disease Burden results from comparing three types of vaccine program strategies

Table 2 - Cumulative Disease Burden for Vietnam Life

	No Vaccine	Vaccine	Under 2 Catchup
Years of life lost due to premature mortality	318,390	65,265	55,378
Years of life lost due to disability	14,115	2,765	2395
Total Disability Adjusted Life Years	332,505	68,030	57,773

Table 3 - Years Economic Burden Results for Vietnam

	No Vaccine	Vaccine	Under 2 Catchup
Hospitalized	1,700	353	352
Rx outpatients	974,491	203,035	171,029
Rx health centers	6,671	1,390	1,170
Rx sequelae	634	117	104

It can be seen that the results for a catch-up program that includes all children under 2 years of age will be the most effective in reducing total numbers of cases and numbers of deaths. We have run the program for a large number of data sets from many countries which are resource-poor or emerging from this category, such as in the case of Vietnam. Results summarizing the decreases in % of cases and deaths for the most vaccine vs. the most effective program of under 2 catch-up are shown in Table 1. The dramatic decrease in mortality and disability for the two options is illustrated in Table 2. And, the impact on costs for hospitalization and treatment of disabilities is just as dramatic as shown in Table 3 on economic burden.

The above is just one example out of many showing the results of the model, for which assumptions such as protective herd effects or serotype replacement can be adjusted. We next show – in Figure 4 – a graph of how the percentage of cases averted and the percentage deaths averted will show considerable improvement over the early years of the vaccine program, and then level-off towards the end of a ten year period as most children in a population have been reached by the program. Figure 6 shows the dramatic effect of decreased numbers of expected cases of mental retardation from meningitis that could be averted by the introduction of the PCV immunization as administered to newborns, or with catch-up programs for immunizing children under-1 and under-2 for two levels of assumed efficacy.

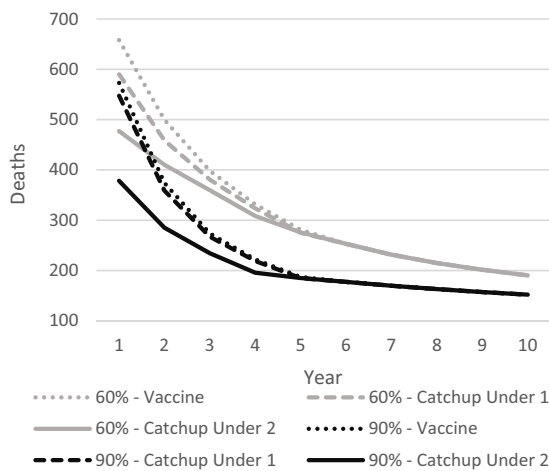


Figure 4 - Deaths averted in Vietnam for 60% and 90% vaccine efficacy rate

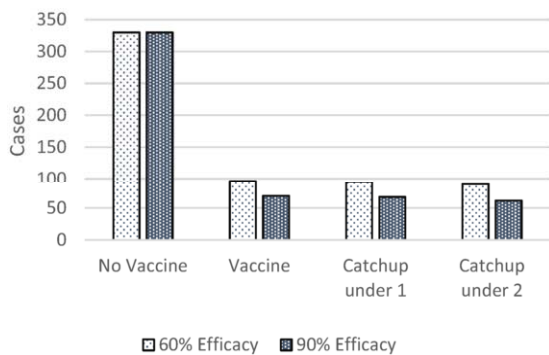


Figure 6 - Cases of Retardation for Vietnam

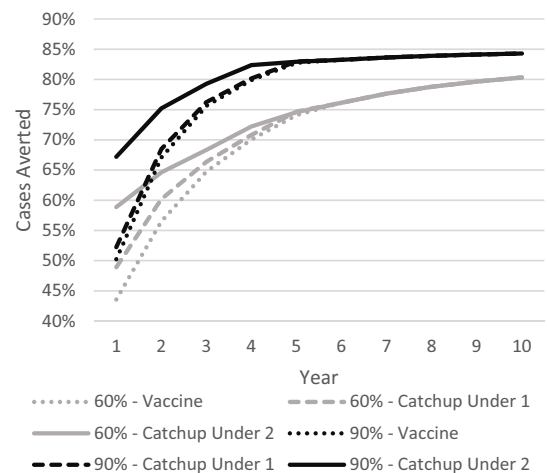


Figure 5 - Cumulative Percentages of Cases and Deaths Averted as a vaccination program progresses over 10 years. Including 60% and 90% vaccine efficacy rate

parameters to carry out sensitivity analyses, and ease of use of the software we have described show that it represents a practical software tool to support policy-makers in testing the possible outcomes of their choices of vaccination program strategy.

Acknowledgments

The authors thank their collaborators Dr. Frank Sonnenberg and, with especial gratitude Dr. C. Greg Hagerty, who was instrumental in producing the first implementation of this system and has since passed away, for their work on the earlier version of the model and software.

Discussion

The results reported here show how the efficient and effective model of reductions in disease burden and costs for a pneumococcal conjugate vaccination program can play out over the course of the program. The tradeoffs between different types of assumptions about population coverage and vaccine effectiveness are the most important practical results which can be derived from these types of analyses in order to assist policy makers in choices of program strategy to adopt given the resource constraints that they face.

Conclusion

The description of the model and implementation with results for an example country showed how the impact of PVC programs with catch-up for children up to 2 years of age can be considerable, with cases and deaths averted amounting to over 70 to 80% of those that would occur without a vaccination program. The portability, ease of changing

References

- [1] C.I. Siettos and L. Russo, Mathematical modeling of infectious disease dynamics, *Virulence* 4 (2013), 295-306.
- [2] A. Clark, B. Jauregui, U. Griffiths, C.B. Janusz, B. Bolanos-Sierra, R. Hajjeh, J.K. Andrus, and C. Sanderson, TRIVAC decision-support model for evaluating the cost-effectiveness of Haemophilus influenzae type b, pneumococcal and rotavirus vaccination, *Vaccine* 31 (2013), C19-C29.
- [3] S. Clark, A. Cohen, D. Constenla, F. Sonnenberg, C. Kulikowski, and A. Sinha, Estimating the impact of pneumococcal conjugate vaccines in resource-poor settings: a portable health policy model for new vaccine introduction, Proceedings SMDM, Miami, FL, 2014.
- [4] L.J. Wolfson, F. Gasse, S.P. Lee-Martin, P. Lydon, A. Magan, A. Tibouti, B. Johns, R. Hutubessy, P. Salama, and J.M. Okwo-Bele, Estimating the costs of achieving the WHO-UNICEF global immunization vision and strategy, 2006-2015, *Bull World Health Organ* 86(1) (2008), 27-39.
- [5] B. Temple, U.K. Griffiths, E.K. Mulholland, F.T. Ratu, L. Tikoduadua, and F.M. Russell, The cost of outpatient pneumonia in children <5 years of age in Fiji, *Trop Med Int Health* 17(2) (2012), 197-203.
- [6] A.D. Clark, U.K. Griffiths, S.S. Abbas, K.D. Rao, L. Privor-Dumm, R. Hajjeh, H. Johnson, C. Sanderson, and M. Santosham, Impact and cost-effectiveness of Haemophilus influenzae type b conjugate vaccination in India, *J Pediatr* 163(1 Suppl) (2013), S60-72.

- [7] U.K. Griffiths, A. Clark, and R. Hajjeh, Cost-effectiveness of Haemophilus influenzae type b conjugate vaccine in low- and middle-income countries: regional analysis and assessment of major determinants, *J Pediatr* **163**(1 Suppl) (2013), S50-S59.e9.
- [8] D.D. Anh, A. Riewpaiboon, H. Tho le, S.A. Kim, B. Nyambat, and P. Kilgore, Treatment costs of pneumonia, meningitis, sepsis, and other diseases among hospitalized children in Vietnam, *J Health Popul Nutr* **28**(5) (2010), 436-442.
- [9] M.G. Lucero, V.E. Dulalia, L.T. Nillos, G. Williams, R.A. Parreño, H. Nohynek, I.D. Riley, and H. Makela, Pneumococcal conjugate vaccines for preventing vaccine-type invasive pneumococcal disease and x-ray defined pneumonia in children less than two years of age, *Cochrane Database Syst Rev* **4** (2009), CD004977
- [10] Kenya SIR model (Flasche S, Edmunds J. 2014. London School of Hygiene and Tropical Medicine, unpublished – personal communication); Expert opinion (Johns Hopkins International Vaccine Access Center)
- [11] S. Flasche, J.A. Scott, W.J. Edmunds, A. Sinha, A. Cohen, F. Sonnenberg, C. Kulikowski, and D. Constenla, Estimating the cost-effectiveness of catch-up vaccination with PCV in GAVI-eligible countries, Technical Report International Vaccine Access Center (IVAC) Johns Hopkins Bloomberg School of Public Health, February 28, 2014.

Address for correspondence

Casimir A. Kulikowski,
 Department of Computer Science, Hill Center, Busch Campus.
 Rutgers-The State University of New Jersey. Piscataway, New Jersey
 08855, USA.
 Email: kulikows@cs.rutgers.edu (preferred).
 Phone: 732-277-3803. Fax: 732-445-0537

Enabling Patient Control of Personal Electronic Health Records Through Distributed Ledger Technology

James Cunningham and John Ainsworth

The Health eResearch Centre and The Farr Institute, Division of Informatics, Imaging and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester

Abstract

The rise of distributed ledger technology, initiated and exemplified by the Bitcoin blockchain, is having an increasing impact on information technology environments in which there is an emphasis on trust and security. Management of electronic health records, where both conformation to legislative regulations and maintenance of public trust are paramount, is an area where the impact of these new technologies may be particularly beneficial. We present a system that enables fine-grained personalized control of third-party access to patients' electronic health records, allowing individuals to specify when and how their records are accessed for research purposes. The use of the smart contract based Ethereum blockchain technology to implement this system allows it to operate in a verifiably secure, trustless, and openly auditable environment, features crucial to health information systems moving forward.

Keywords:

Electronic Health Records; Health Information Systems; Technology

Introduction

Medical data, both routinely collected and specifically studied, is increasingly being recorded, represented, and stored electronically [1]. Having access to these resources in electronic form is immensely beneficial to researchers, allowing novel research methods to be applied over large volumes of data in a way that would not have been possible even a decade ago [2, 3]. This switch to the digital does, however, bring with it problems relating to the physical and ethical security of medical data [4]. Data stored electronically is much easier to copy, distribute, and mine for confidential information. Breaches of security and the loss or misappropriation of data negatively impact the public perception of medical research and threaten to bring down regulatory restrictions that will prevent or hinder future research [5]. The current state of information security within the medical informatics domain makes controlling and identifying such breaches difficult [6].

Further, there is a gap between the ideal needs of the research community and ethico-legal restrictions on the use of personal medical record data [7]. The research community's desire for essentially unfettered access to data is checked by the legal responsibility of data owners to guarantee the privacy of the patients whose data they protect. There is also a recognized need for public involvement in the research process [8] and even where patient data is de-identified and used in the aggregate, granting of explicit consent for data use has been identified as the preferred model for half of all patients, with increased awareness of Electronic Health Records (EHRs) impacting positively on willingness to consent to research [9]. The best way of achieving and maintaining this balance

between trust, security and admittance of public participation is an open question [10] — how can medical data be shared in a way which still, at its core, guarantees as far as possible the security and integrity of the data being shared? Addressing these challenges will be of crucial importance to health informatics moving forwards.

We present a proof-of-concept system for enabling fine-grained specification of access control policies as pertaining to third-party access to electronic health records on an individual level, which goes some of the way towards tackling these issues. Giving patients control of access to their own records whilst giving research organizations, possibly from the private commercial sector, the ability to directly reach out to patients and request access to medical data, opens up a series of issues around trust and security. We have implemented this system using the Ethereum platform, a modern, smart contract based, distributed ledger system [11]. This choice of platform not only allows for a natural expression of a solution to the problem we address but also explicitly addresses and solves the underlying complex issues of trust and security.

Methods

Enabling direct patient involvement in controlling the use of medical data, and doing so in open and secure manner will enhance both uptake and acceptance of medical informatics platforms aimed at enabling access to research data. We aimed to develop a core Application Programming Interface (API), for use as a service within a wider platform, that would enable a permission system through which patients could both specify who could access their records and to review the uses to which their data have been put. We identified this as the most fundamental building block of any system that would purport to enhance patient control in a research-oriented informatics system.

The requirements for the design of the API were gathered through an analysis existing systems for sharing and reuse of medical data deployed within the North of England [12]. This gave us an overview of the core functionality required by such a system.

The design of the API proceeded first with the identification of the key class of actors who would interact with the system. A series of core requirements based around ensuring requisite levels of security, trust environment, and transparency were then developed. Following this development, a series of use cases were produced that specified the ways in which the system actors would be able to interact with the API in order to achieve the overarching goal of enabling patient control of access requests to EHR data.

In implementing the system to address these use cases we chose distributed ledger technology ('DLT') as the underlying technical implementation. DLT is the mechanism, in terms of

data structures and associated computational methods, underlying the Bitcoin cryptocurrency, in which the specific instance of DLT is the Blockchain [13]. The driving use case behind the design of the Blockchain was the desire to allow for a decentralised transaction ledger; the provision of a canonical global account balance for all holders of the Bitcoin currency, without reliance on a trusted third party as a point of control [14]. The Blockchain implementation further provides both a public key-based infrastructure for account identification and control, and a 'mining' mechanism (a computational competition which both incentivizes hosting of the peer-to-peer Bitcoin network and solves the so-called 'Byzantine Generals problem' — that of guaranteeing consensus amongst distributed network nodes containing potential bad actors) [15]. Since its conceptualisation DLT found extended use across a range of application areas [16] and has been extended with the addition of 'smart contracts', a means of adding distributed computational processing to the underlying transactional ledger [17].

The core features of DLT were identified to meet the specified system requirements. Specifically, we chose the Ethereum Platform [11] to host the API given an assessment of its capabilities in producing a system to fulfil the outlined use cases. Ethereum itself provided both the core DLT blockchain capabilities allowing for hosting within a trustless, secure environment, and a smart-contract implementation allowing for the programmatic implementation of the API directly on the platform itself.

Results

Presented in the following section are the results of our design and implementation in terms of the actors we identified, the core requirements, the driving use cases and a specification of the API.

The core actors within the system were identified as follows:

- **Public Participant** — someone for whom associated EHRs are accessible within some system who will be granted control of a permission system for accessing those records.
- **Research Organization** — Representing an organizational entity that wishes to request access to EHR data for research purposes.
- **Data Custodian** — An organizational entity with ultimate control of source electronic medical data

We identified the following underlying requirements for the system:

- **Trustless** — Reliance on a third party to maintain control of some or all of the system functionality decreases trust and adds a single point of failure to the system.
- **Incentivized** — With use of a distributed technology, it is required that participants within the system should share the burden of hosting the system itself. A mechanism is needed to ensure this happens.
- **Secure** — The system needs to be secure in that it must prevent actions being performed by entities or actors not specified as being allowed to perform those actions.
- **Identifiable** — Actors need to be strongly and verifiably identifiable within the system. Recognizing that key management by the lay public is a difficult issue, formal identity management may be delegated to other actors (for example Data Custodians

performing actions within the system on behalf of users)

- **Transparent** — Perceived trust in the system is crucial so all actions that take place within the system need to be publicly auditable.

The following use cases were identified as the drivers of the API design:

- A **Research Organization** can publish a request for data in the form of a **Research Proposal**. The **Research Proposal** will outline what medical data is required for the research, the form of that data and the limits of its use.
- A **Data Custodian** can vet a **Research Proposal** and decide whether it is published to **Public Participants** whose data that custodian safeguards.
- A **Public Participant** can set a general preference for how their medical data should be used. This will allow options for allowing or denying all requests, or granting permissions on a proposal-by-proposal basis.
- A **Public Participant** can view lists of **Research Proposals** which would utilize their private medical data.
- A **Public Participant** can specify an option against a **Research Proposal** indicating whether or not they will grant permission for the use of their data within that proposal.
- A **Data Custodian** can request a list of patients who have consented for a particular **Research Proposal**

We chose to implement the code for enacting these use cases using DLT. Through creating a distributed ledger instance, records of what proposals have been offered and which participants have granted, either implicitly through global options, or explicitly against a particular proposal, are recorded within the 'blocks' that form the ledger. These use cases were then implemented directly in the form of smart contracts deployed within the blockchain instance. A smart contract is an executable piece of code that references the current state of the ledger and can write back to it. Smart contracts are authored in Solidity, a Turing complete programming language, broadly similar in syntax and structure to javascript. These were compiled into Ethereum compatible byte-code and hosted on the ledger instance. An illustrative extract of such a smart contract is given below.

```
contract Participant{
    address admin;

    function Participant (address owner_address) public {
        owner = owner_address;
    }

    function Owner () constant returns (bool confirmation){
        return msg.sender == owner;
    }

    enum Option { Grant, Deny }

    struct Participant {
        uint identifier; // every participant has an id number
        address participant_addr;
        address id_custodian; // the id of this patient's corresponding data custodian
        string record;
        uint status;
        uint[] institution;
    }

    ...
}
```

Figure 1— A Portion of a Smart Contract

Accounts for actors within the system (patients, research organizations and data custodians) were either created directly as user accounts within the Ethereum system (with associated public/private key pairs), or, in the case of patient accounts,

managed via a data custodian account -- that is, public participants controlled their accounts via functions exposed by the data custodian entity. This mitigated the difficulty of requiring public participants in the system to manage their own public/private key pairs.

The hosting environment for the API consisted of a private instantiation of the Ethereum platform -- one entirely distinct from the canonical, publicly accessible peer-to-peer instance. Further, node hosts were set up inside in virtual machines and firewall rules put in place to restrict peer-to-peer connections to other known hosts. Access to the system in its entirety is then ultimately controlled through network-level security.

Discussion

We chose distributed ledger technology as the underlying technical implementation for this service as it provided, as an integral part of the technology itself, a secure framework that enables deployment within a network not reliant on a central point of trust. As stated, establishing a balance between the desires of the research community, the ethical and legal obligations of data custodians, and the ultimately decisive needs as well as the perceptions of the public is of key importance in establishing sustainable medical informatics frameworks enabling the use of electronic health data. The nature of distributed ledger technology, in that it provides an auditable, accountable framework removed from a single point of trust (and hence possible failure or compromise), makes it a natural fit for such systems. Identity management is a fundamental requirement of any such system -- inability to establish and ensure identity across a system immediately invalidates any claim to security. Again DLT is built around the concept of identity and public key management, providing a base on which to build systems where identity is a crucial component. Smart contract functionality is provided by the Ethereum platform.

The blockchain data structure is designed for redundancy across a network of peers and smart contract mechanisms deliberately replicate computation of programmatic structures across all nodes in the network. Whilst this provides benefit, as mentioned, in the areas of trust and security it does introduce computational inefficiencies. These raise potential issues in terms of the use of the technology in a real-world setting, particularly one at scale. However, the use cases that the design of the system was based upon do not require absolute real-time instantiation across the system. The preferences set by users and the requests for research data do not need immediate processing or acting on by external services. As such this mitigates to a degree the natural inefficiencies in the system. The reliance of current distributed ledger technology on essentially arbitrary computation (token mining) again introduces inefficiencies into the system. The original aim of the mining mechanism in the Bitcoin scheme -- that of incentivizing distributed participation and as a means of introducing and distributed currency -- does not translate entirely into an environment with a higher degree of trust between participants. Such mechanisms can still be leveraged as a means of ensuring fair participation in the ecosystem; for example with tokens, rather than providing economic value, it can be used as a proof-of-participation and required for ongoing use of the system.

Distributed ledger technology, in particular Ethereum, are new and hence relatively untested technologies. Whilst their primary use case as financial tools with intrinsic economic value incentivizes both secure implementation and testing through real world exposure, bugs and security flaws are do persist. Further, the rapid pace of innovation and evolution of

the platforms will carry increased risks of potentially harmful design flaws becoming apparent.

Constant evaluation and auditing of the underlying technologies in terms of monitoring and addressing security flaws will therefore be crucial in the future use of this technology. The current system we have developed exists as a stand-alone API with the explicit design goal of acting as a component in a wider service oriented medical informatics platform. Future work will address this as we aim to roll out a test deployment of the system within an infrastructure for enabling actual access to data within a real-world setting. Also, given the potential issues surrounding public perception of distributed ledger technology, both in respect to it being a nascent and relatively untested technology, and its association with negative reporting and use within black markets, work will be done on gauging public attitude towards the use of the technology within a healthcare environment.

For any medical informatics system the key driver of its implementation, and the ultimate gauge of its success, will be measured in the benefit it brings to medical care and practice. In moving forwards the state of the art in terms of security, privacy and accountability, distributed ledger technology has the potential to add a significant degree of trust to the functionality of medical informatics systems. With this enhanced trust will come the ability to utilize medical data within richer settings and with a wider range of participants -- something that will ultimately improve medical care and practice moving forwards.

Conclusion

We have identified a series of key requirements for enabling enhanced patient control of EHRs and have developed a proof-of-concept API meeting these requirements. Distributed ledger technology was chosen specifically to meet what we see as the fundamental issues in health informatics, namely trust and security. Whilst problems do exist, particularly with respect to the fundamental inefficiency and current immaturity of the technology, distributed ledger technologies offer unique solutions in the health informatics domain and will inevitably see increasing use in the field in future.

References

- [1] A. K. Jha, D. Doolan, D. Grandt, T. Scott, and D. W. Bates, "The use of health information technology in seven nations," *International Journal of Medical Informatics*, vol. 77, no. 12, pp. 848 -- 854, 2008.
- [2] T. Murdoch and A. Detsky, "The inevitable application of big data to health care," *JAMA*, vol. 309, no. 13, pp. 1351--1352, 2013.
- [3] R. Bellazzi, "Big data and biomedical informatics: a challenging opportunity," *Yearbook of medical informatics*, vol. 9, no. 1, p. 8, 2014.
- [4] M. Wiesenauer, C. Johner, and R. Rohrig, "Secondary use of clinical data in healthcare providers-an overview on research, regulatory and ethical requirements," *Stud Health Technol Inform*, vol. 180, pp. 614--8, 2012.
- [5] H. K. Patil and R. Seshadri, "Big data security and privacy issues in healthcare," in *2014 IEEE international congress on big data*, pp. 762--765, IEEE, 2014.
- [6] T. Sahama, L. Simpson, and B. Lane, "Security and privacy in health: Is it possible?," in *e-Health Networking, Applications & Services (Healthcom), 2013 IEEE 15th International Conference on*, pp. 249--253, IEEE, 2013.
- [7] G. Chatellier, V. Varlet, C. Blachier-Poisson, *et al.*, "big data and open data: What kind of access should researchers enjoy?," *The rapie*, vol. 71, no. 1, pp. 107--114, 2016.
- [8] L. Doyal, "Informed consent in medical research. journals should not publish research to which patients have not given fully informed consent--with three exceptions.," *BMJ: British Medical Journal*, vol. 314, no. 7087, p. 1107, 1997.
- [9] F. Riordan, C. Papoutsis, J. E. Reed, C. Marston, D. Bell, and A. Majeed, "Patient and public attitudes towards informed consent models and levels of awareness of electronic health records in the UK," *International Journal of Medical Informatics*, vol. 84, no. 4, pp. 237 -- 247, 2015.

- [10] N. C. Lea, J. Nicholls, C. Dobbs, N. Sethi, J. Cunningham, J. Ainsworth, M. Heaven, T. Peacock, A. Peacock, K. Jones, *et al.*, "Data safe havens and trust: Toward a common understanding of trusted research platforms for governing secure and ethical health research," *JMIR Medical Informatics*, vol. 4, no. 2, 2016.
- [11] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, 2014.
- [12] J. Ainsworth and I. Buchan. "Combining health data uses to ignite health system learning." *Methods Inf Med* 54,6 pp. 479-87, 2015.
- [13] R. Ali, J. Barrdear, R. Clews, and J. Southgate, "Innovations in payment technologies and the emergence of digital currencies," *Bank of England Quarterly Bulletin*, p. Q3, 2014.
- [14] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system." <https://bitcoin.org/bitcoin.pdf>, 2008. [Online; accessed 28Nov-2016].
- [15] A. Miller and J. J. LaViola Jr, "Anonymous byzantine consensus from moderately-hard puzzles: A model for bitcoin." <http://nakamotoinstitute.org/research/anonymous-byzantine-consensus>, 2014. [Online; accessed 28-Nov-2016].
- [16] M. Pilkington, "Blockchain technology: principles and applications," *Research Handbook on Digital Transformations*, edited by F. Xavier Oleros and Majlinda Zhegu. Edward Elgar, 2016.
- [17] S. Omohundro, "Cryptocurrencies, smart contracts, and artificial intelligence," *AI matters*, vol. 1, no. 2, pp. 19–21, 2014.

Address for correspondence

James Cunningham, james.a.cunningham@manchester.ac.uk

Enhancing Children's Safety by Barcoding Implementation at Breast Milk Feeding

Mariana Daus, Thelma Maydana, Daniel Rizzato Ledé, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

When newborns remain hospitalized in a neonatal intensive care unit, they are often unable to feed themselves and receive human milk through enteral nutrition devices such as orogastric or nasogastric probes. Therefore, the Nursing staff is responsible for the fractionation, storage and administration of human milk. Breast milk has a great biological complexity being the optimal food for the baby to provide all the nutrients needed. At the same time, it is a bodily fluid that carries the risk of disease transmission if not administered properly. Patient safety should be a priority in healthcare, and health information technologies could be used to avoid preventable adverse events. Barcoding technology has the ability to accurately verify patient identity and prescription accuracy before milk administration. This paper describes the steps followed to implement breast milk barcoding technology in an academic tertiary hospital.

Keywords:

Breast Feeding; Barcoding; Patient Safety

Introduction

The Hospital Italiano de Buenos Aires has been engaged in quality and safety improvements for a long time. Currently, we are working to implement the barcoding verification technology in the circuit of human milk.

Human milk is the first natural food of newborns, and it provides all the energy and nutrients they need during their first months of life. Further on, it continues to provide at least half of their nutritional needs during the second half of the first year and up to a third during the second year of life. It promotes sensory and cognitive development, and protects the baby from infectious and chronic diseases. It has proven to reduce child mortality by preventing childhood diseases, such as diarrhea or pneumonia, and even favors recovery in case of illness [1].

When a baby is premature, ill, or away from her mother for any reason, it is very important to promote breastmilk manual or mechanical extraction for baby feeding. In a hospital setting, health professionals are responsible for the babies' feeding if they can not breastfeed naturally. Preterm newborns before week 32 do not have an effective suction, so enteral feeding is done by orogastric probe with gavage technique.

Obtaining the human milk requires breastpumps and sterile circuits with aseptic technique performed by trained mothers, ideally monitored by specialized staff [2].

In the Neonatal Intensive Care Unit (NICU), the nursing staff is responsible for the fractionation, conservation and administration of the human milk. At every step, it is essential to verify the patient identity for safety and quality of care. The

"5 rights" rule recalls each element that should be verified before administration: the right patient, the right medicine, the right dose, the right path, at the right time. Several strategies are constantly designed to reduce the incidence of errors in healthcare. A great deal of support has been found in information technology, which has already shown enormous benefits for medicine. Barcode is a simple but very useful technology inherited from commercial industry. Its implementation and universalization in healthcare has been endorsed to check the patient information for their safety. This simple system allows quick, safe and error-free access to meaningful information required in diagnostic tests or during treatment administrations [3].

The objective of the present paper is to describe the former process of the breast milk circuit in our hospital, and the subsequent steps taken to design and implement the barcode verification technology.

Methods

Setting

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit health care academic center founded in 1853, with over 2,700 physicians, 2,700 other health team members (including 1,200 nurses) and 1,800 administrative and support employees. The HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 homecare beds, 25 outpatient clinics, and 150 associated private practices located in Buenos Aires city and its suburban area. It has a Health Maintenance Organization (Plan de Salud) that covers more than 150,000 people and also provides health services to another 1,500,000 people who are covered by affiliated insurers. Between 2013 and 2014, over 45,000 inpatients were admitted to its hospitals. There were 45,000 surgical procedures (50% ambulatory) and 3,000,000 outpatient visits. In addition, the HIBA is a teaching hospital, with over 30 medical residency-training programs and 34 fellowship programs. There are currently 400 residents and fellows in training.

Since 1998, the HIBA runs an in-house developed health information system, which includes clinical and administrative data [4]. Its Electronic Health Record (EHR) system called Italica, is an integrated, modular, problem oriented, and patient centered system that works in different clinical settings (outpatient, inpatient, emergency and home care). Italica allows computer physician order entry for medications, medical tests, and storage and retrieval of tests results. It also includes images through a picture archiving and communication system. It has been recently certified by the HIMSS as level 6+ in the Electronic Medical Record Adoption Model (EMRAM), being the first hospital in Argentina and the second in Latin America reaching this stage [5].

The HIBA achieved Joint Commission International safety and quality certification in 2015 [6]. One major addressed issue was the right identification of patients. To this end, in 2010 the Department of Health Informatics developed and implemented an inpatient bracelet identification system with barcoding technology. This code can be scanned to verify patient identity before every intervention (patient transfer, medication administration or other procedures).

Nowadays, the process of scanning the tag of every medication before its administration is performed just in the Adult Intensive Care Unit. In order to certificate EMRAM level 7, we should perform barcoding verification in every inpatient setting, for each substance administration (including breast milk, human blood products and medications).

The neonatal intensive care unit (NICU) is divided into three different areas: the Intensive Care Unit (22 beds), the Hospitalization, and the Follow-up Clinic for premature and high-risk newborns (25 beds).

Direct mother breastfeeding does not need patient-substance verification, but sometimes the hospitalized newborns are not able to feed by suction and need other route for enteral nutrition, thus requiring special assistance from nurses [7]. Thus, we focused only in the intensive care sector where the circuit of breast milk extraction, storage and administration takes place.

Methodology

We carried out an observational analysis of the processes, supported by bibliographical evidence. We assessed the areas of lactarium (lactation room), the milk fractionation room and the neonatal intensive care room (NICU) by reviewing the whole circuit: first the extraction of breast milk, storage, sterile fractionation and finally the milk administration in the patient unit. This revision work was guided with project management methodology [8]. Each process was plotted with flowcharts.

Results

Process traceability

We plotted the milk process in a flowchart using all the information collected during the study. Afterwards, we drew a

journey map highlighting risk and opportunities of each step (Figure 1).

Breastmilk feeding prescription

The circuit starts with the feeding prescription. Pediatricians fill the milk order through a CPOE, but in plain text (not structured) including heterogeneous data. As these prescriptions are not linked to the medication label printer, nurses must read the instructions and label the milk manually.

Lactation room

The first observed place was the lactarium. It is a room specially equipped for the extraction and storage of human milk, where a breastfeeding woman can use a breast pump in private. It has 3 electric pumps and a refrigerator. Each mother uses a breast pump kit (with sterilized materials) and then discards it for sterilization. This work is supervised by specialized staffs, which coordinate the activities of the sector and educate each mother to correctly perform the extraction technique. They are responsible for printing patient identification labels from the EHR containing the baby's hospitalization data. These labels are stored in separate plastic folios in a common folder located inside the lactarium.

Staffs verify the newborn feeding prescription, and deliver each patient's label to the corresponding mother to identify their milk every time it is extracted.

The women put their milk in sterile flasks labeled with the babies' identity and then handwrite date and time, before storing them in the refrigerator.

Milk fractionation room

This area is intended for sterile manipulation and fractionation of the milk. It is of vital importance, since it prevents potential risk of cross-infections. [9] Three nurses are trained and assigned to this process.

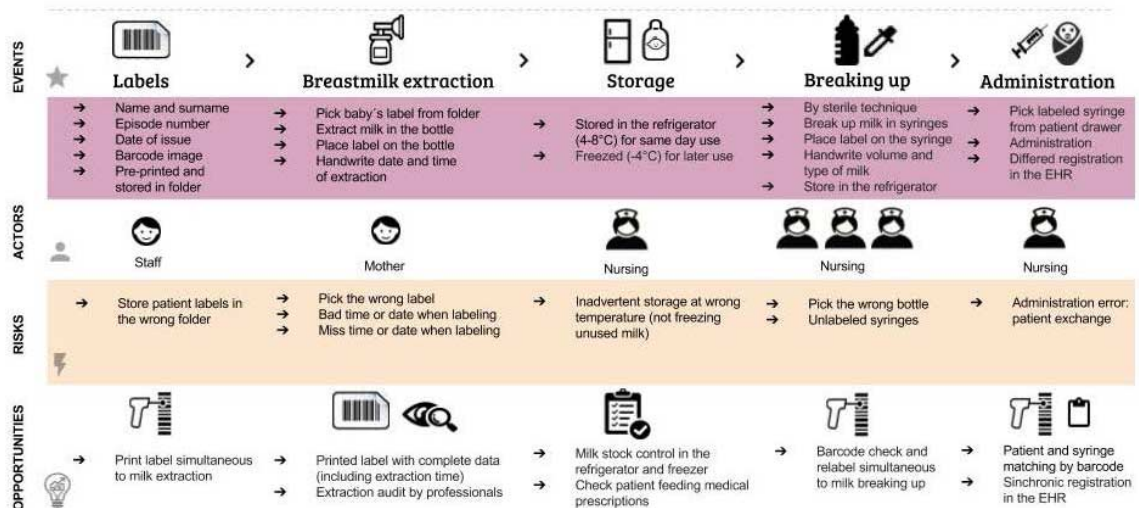


Figure 1 - Process journey map

Initially, the newborn feeding prescription is checked for each patient in the EHR through a PC located in the room. At the moment, it is not possible to print this type of prescriptions in specific labels. Therefore, nurses use generic labels with patient ID to identify the milk syringes.

One nurse brings the bottles from the lactarium's refrigerator. The milk from each bottle is divided into syringes following the prescribed dose. Afterwards, the syringes are labeled and stored in the corresponding patient's drawer, located in a special refrigerator of this sector.

If any of the milk bottles is not going to be used shortly after, the nurses store it in a freezer to extend the expiration date.

Neonatal intensive care unit (NICU)

There are nine nurses in the neonatal intensive care unit; i.e. two nurses per patient.

The nurses must pick the milk syringe from the refrigerator and go to the patient bedside. Every time an intervention takes place, it is necessary to validate the patient identity by visual confirmation of the incubator label and the patient bracelet. After checking the five rights, they administer the milk to the patient through the corresponding device. Finally, the administration should be recorded in the EHR.

Proposal for improvements in breastmilk prescription

CPOE for pediatricians requires a software adaption to include a specific module for breastmilk feeding prescriptions.

We searched the data warehouse for general orders containing milk prescriptions of the last 6 months to look for examples and patterns. We defined the minimum fields required to structure the posology: type of milk (human), dose, route, infusion method, infusion speed, and frequency.

We sketched initial pen and paper prototypes and validated them with final users. Afterwards we made high fidelity prototypes emulating the interface for medication prescription (Figure 2).

Proposal for improvement in the Lactorium

There are risks of human error in the current process:

1. Asynchronous label printing (long time ahead of extraction) hinder direct patient identity validation.
2. Storage of pre printed labels in plastic folios can lead to wrong bottle attachment.
3. The handwritten extraction date and time can be erroneous or missed.

Therefore, to fulfill the first picking verification it is suggested to corroborate the mother and child's identity. It requires the following modifications:

- **Hardware:** to place a PC with a label printer and a barcode scanner in the Lactorium.
- **Software:** to develop a tool to verify the mother ID by barcode scanning. When the system validates her relationship with the baby, then it prints a label for each bottle of milk. The mother must carry an identification wristband that should be printed in advance at Admission office.

The mother should read the wristband barcode with a scanner through a simple interface of the software (as shown in Figure 3). If the system validates the identity, it would print a label synchronously just prior to milk extraction, adding a timestamp. We suggested including a calculated expiration date, according to storage temperature. [10]

Proposed improvement in the Milk Fractionation Room

The nursing staff is responsible for the storage and fractionation of the milk. There are also risks of human error in the current process:

1. **Fractionation:** staff may confuse the EHR feeding prescriptions or even crossing over with another patient, as they check them just by observation.
2. **Disposal of surplus milk:** the bottle can miss the timestamp, thus preventing the staff to store it in the freezer, as they can not calculate expiration date.
3. **Labeling of prepared syringes:** staffs may place the wrong label or even miss it.

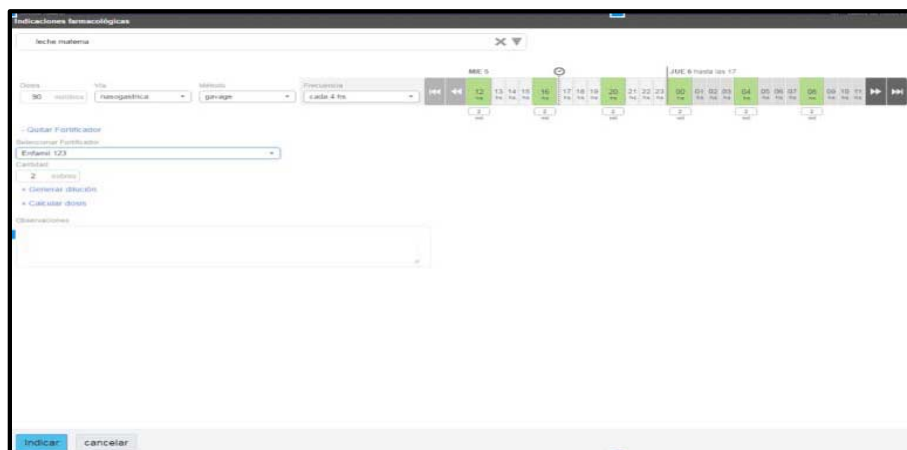


Figure 2 - Human milk prescription prototype.

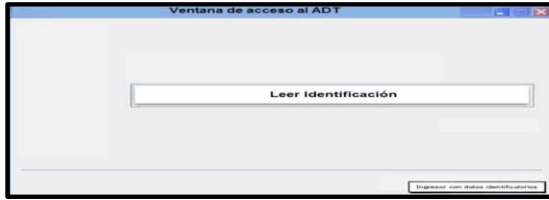


Figure 3 - Identity validation system. After the mother read her wristband with the barcode scanner, it prints a label with the information of the baby and a time stamp.

Therefore, a second picking validation is necessary to corroborate that the right prescription is prepared with the right sample of milk. It would require the following modifications:

- **Hardware:** adapting the existing PC, adding a barcode scanner and a label printer.
- **Software:**
 - Adapting the nurse preparation worklist from the EHR (just intended for medications) to incorporate breast milk prescription.
 - When selecting the prescription to prepare, if the scanned bottle barcode is from the right patient, the system will print a new label containing a detailed description (patient name and ID, type and amount of milk, route and frequency) and a new datamatrix code.

Proposed improvement in the NICU

Prior to milk administration, a third verification is needed to check the right patient identity and the other four rights of the prescription (milk sample, dose, route, time) by barcoding. This would require the following modifications:

- **Hardware:** incorporation of a barcode scanner in each patient unit (attached to the bedside device: tablet or computer).
- **Software:** adaption of the nurse module of the EHR to allow barcode synchronous verification before milk administration in the patient unit.

Discussion

We intended to design and implement a reliable system for closed loop breast milk circuit. Barcoding technology can enhance patient safety by verifying the identification at the time of the removal, storage, fractionation and administration of human milk, and establishing a physical barrier that can decrease the errors due to human factors in every stage of the circuit [11].

We started our project defining objectives, tasks, desired scope, and work equipment. During the research it was possible to define the actors, the events, the risk, and opportunities for improvement. We could analyze the best way to achieve a rate of verification by barcoding greater than 95%.

The incorrect administration of breast milk involves several complications: may cause the transmission of diseases, causing additional examinations to the patient and the loss of confidence of their relatives in the health professionals or the institution.

The manual verification processes of the five rights for medication administration are prone to human error. The barcoding of breast milk provides a simple technological solution to improve safety in the preparation and management of the product because it acts as a barrier that can prevent adverse events. The barcode is read by an optical scanner, which is routed directly to an active computer application at the time of reading.

The most important negative risk would be the resistance to change of the staff involved in the implementation of new tasks. It is of the utmost importance to create solution strategies before the events happen and promote the participatory change, making available the new knowledge to the individuals or groups involved [12].

The successful implementation of the barcode in the human milk will be affected by the improvement of the process before they start the automation, overcoming the barriers such as the time and expense of training, the costs of software and hardware, the flow of work and avoid temporary solutions that can weaken the security of the processes [13].

The goal is to apply the existing knowledge and validate it in the design of procedures that help to satisfy a need. The power plan in advance the steps needed to guide the work of the team toward the achievement of the goals, would establish corrections, determine results, and implement new strategies where necessary [14].

Conclusion

In this work we describe the current events and the necessary modifications to achieve the implementation of barcoding in the circuit of extraction, fractionation and administration of human milk, adapted to our local process. This allows the traceability of the milk from the time of removal until the administration. A culture of safety involves users, structural elements, processes, instruments and methodologies based on evidence that minimize the risk of adverse events.

References

- [1] OMS. Lactancia materna. Nutrición. *Salud de la madre, el recién nacido, del niño y del adolescente*. Available from: <http://www.who.int/maternal-child-adolescent/topics/newborn/nutrition/breastfeeding/es/>
- [2] Alimentación homóloga para niños prematuros. Guía temática. Servicio de Neonatología del Hospital Italiano de Buenos Aires 2016. Available from: <https://www1.hospitalitaliano.org.ar/#!/home/neonatalogia/seccion/10895>
- [3] Lopez Gala, B. Tecnologías informáticas en pos de la seguridad del paciente: el código de barras en sanidad. *Revista cubana de Informática médica*. 1 (5), falta N° pag
- [4] Luna D, Plazzotta F, Otero C, González Bernaldo de Quirós F, Baum A, Benítez S. *Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires*. Hospital Italiano de Buenos Aires, CEPAL, 2012.
- [5] First hospital in Argentina awarded for digital maturity n.d. Available from: https://www.eiseverywhere.com/file_uploads/b4ead7f8dcbe30ebefb601e62dfeaa_Argentina_HospitalitalianoPressRelease.pdf
- [6] Noticias Departamento de Informática en Salud del Hospital Italiano. Único Hospital en Argentina en acreditar calidad de Historia Clínica Electrónica 2016. https://www1.hospitalitaliano.org.ar/informad/index.php?contenido=ver_curso.php&id_curso=19752
- [7] Soria R.M. Alimentación enteral en neonatos. *Revista de enfermería neonatal*, 8 (2010), 11-14.
- [8] Available from: <http://www.pmi.org/>
- [9] Steele C, Bixby C. Centralized Breastmilk Handling and Bar Code Scanning Improve Safety and Reduce Breastmilk Administration Errors. *Breastfeed Med*, 9 (2014), 426-429.
- [10] Steele C, Czerwin A, Bixby C. Breast milk bar code scanning results in time savings and staff efficiency. *J Acad Nutr Diet*, 115 (2015), 23-26.
- [11] Fleischman EK. Innovative application of barcoding technology to breast milk administration. *J Perinat Neonatal Nurs*. 27 (2013), 145-150.

- [12] Tobón, S. *Método de trabajo por proyectos*. Madrid, Uninet, 2006.
- [13] Hersey, Paul y Blanchard, Kenneth. *Administración del Comportamiento Organizacional*, Prentice Hall, 1998.
- [14] Dougherty D, Nash A. Bar coding from breast to baby: a comprehensive breast milk management system for the NICU. *Neonatal Netw* 28 (2009), 321–328.

Address for correspondence

mariana.daus@hospitalitaliano.org.ar

thelma.maydana@hospitalitaliano.org.ar

A Mobile System for Music Anamnesis and Receptive Music Therapy in the Personal Home

Kerstin Denecke

Bern University of Applied Science, Bern, Switzerland

Abstract

Receptive music therapy is active hearing of music that is specifically selected to cause a certain effect on a person, such as pain reduction, mental opening, confrontation etc. This active, guided hearing could be helpful as a supporting ritual for patients at home and could extend traditional therapy. However, patients are often unable to select the music pieces that might be helpful for them in a current situation. We are suggesting a self-learning decision support system that allows a patient to answer questions on music anamnesis, is ready for inclusion into an electronic health record, and which enables a therapist to compile a therapeutic music program for the patient at home. Beyond this, the system also suggests appropriate music and duration of listening based on the patient's reported current mental state. In this paper, a concept for such a mobile system for receptive music therapy will be proposed.

Keywords:

Music Therapy; Decision Support Techniques; Mobile Applications

Introduction

Music therapy is defined as “the use of sounds and music within an evolving relationship between patient and therapist to support and encourage physical, mental, social, emotional and spiritual well-being.” [1]. It is a young treatment method indicated in patients with autism [2], dementia [3], depression [4], pain [5], and other conditions. In 2005, it was included in the DRG catalogue for multimodal clinical pain therapy. The objective is primarily to help clients improve their health in several areas, such as cognitive functioning, motor skills, emotional development, social skills, and quality of life. To achieve this, patients are encouraged to experience music by free improvisation, singing, listening, and moving to music. Active and receptive music therapy can be distinguished. We are focussing on receptive music therapy in this paper.

Receptive music therapy comprises “active hearing of selected music pieces under certain conditions” [6]. The difference from everyday listening of music is that certain effects are to be achieved, e.g. a confrontation, spiritual opening etc., which requires a careful selection of the music pieces. In addition, there is a concrete guide which encourages listening.

Even though the therapist-patient relationship is important in this context, there are use cases where it would be helpful for patients and for the treatment – similar to doing physical exercises at home – if aspects of receptive music therapy could be continued in a guided manner in the personal home of a patient. For example, from the treatment of patients with

chronic pain, it is known that patients can be equipped with a therapeutic program for home use, but this requires exact specification of usage to create an individual ritual against pain [7].

The problem is that clients have difficulties in deciding which music would be good for them in a particular situation. They could even select music that could harm them or achieve undesirable results. Existing music compact discs (CDs) for individual music therapy, e.g. by W. Zeitler [6] are described very well with a lot of practical experiences. Nevertheless, Zeitler reports that the help-seeking clients often do not understand these descriptions and would choose some music which does not help them.

A decision support system implemented on a mobile device such as a smartphone or tablet as designed in this paper could help in selecting the appropriate music and make suggestions according to a therapy plan. More specifically, the objective of this paper is to design a system that supports a patient in using the therapeutic capacities of music therapy at home. Taking existing practice into account, the system will in particular support the therapist in collecting music anamnesis and in creating therapeutic music therapy programs. Starting from a requirement analysis, we are developing a concept for a mobile system supporting music therapy.

Using digital media in music therapy has gained interest in the last couple of years. Computers, smartphones, and tablets are mainly used for recording, composing, adapting music pieces and accessing music libraries [8-9]. Several case studies have illustrated multiple uses of technology with young adults [10-11]. In particular, recording technologies are exploited for enabling immediate capturing of spontaneous music making during therapy session or with others. There are apps available that allow users to arrange music into loops (e.g. Relax melodies, <http://www.ipnossoft.com/app/relax-melodies/>) in order to create a personal meditation or relaxation program. However, to the best of our knowledge there is no system available that supports the music anamnesis and therapeutic home usage we are considering.

Material and Methods

We performed a requirement analysis by means of a structured questionnaire. The current and target situations had to be described together with functional and non-functional requirements. The questionnaire was completed by one music therapist. Additionally, we collected requirements from existing literature on music therapy. Afterwards, we designed the architecture for the system and developed first mock-ups. In this section, we are describing the current procedure in music therapy and possible usage scenarios for the system as

received through the requirement analysis. Further, we are summarizing the requirements.

Music Anamnesis and Music Therapy

Traditionally, within receptive music therapy, the therapist selects music and plays the recorded music to the patient. Afterwards, a discussion on the feelings caused or other effects takes place. A main characteristic is that selection of the music focuses on symptoms of the patient and his/her personality. Three main issues determine selection of music within the therapy session: the music anamneses which are habits and preferences of music listening of the patient, the current situation (emotional or health state) and the therapeutic goal. We describe these three aspects in more detail below.

In the first session, the music therapist makes a *music anamnesis*. For this purpose, questions are directed to collect information on the patient's experiences with music in childhood and current life, preferences in music listening, etc. (see Table 1). Additionally, emotions caused by music listening will be identified and the patient will be asked to judge personal experiences with music. Further, it is important to determine habits in music listening and preferences towards certain instruments, genres or music pieces. The objective is not to identify the preferred music – this can cause the person to experience past emotions that can become problematic in the current situation. The objective of the music anamnesis is to get an idea of the categories of music that could help a person in situations targeted by the therapy.

Table 1 – Music anamnesis: Examples for relevant aspects

Early experiences with music
- Lullabies, Christmas songs
- Learned an instrument? Which?
Experiences as youth
- Listening with friends
- Party
Habits of music consumption before getting sick
- Situations when music was recognized with positive impacts
Current habits of music consumption
- Kind of music
- Frequency of listening
- When / Situations
When do you listen to music?
- While working
- Dancing
- Learning
- Car driving
Recognized effects or emotions while listening to music
- Getting calm
- Making someone cry
- Tears

For the *current situation of a patient*, his or her mental state is assessed by asking for current feelings and symptoms. Consider for example a patient suffering from chronic pain. The current situation comprises that pain at the given time and the general mental constitution (e.g. feeling entirely surrendered to the pain, depressed, afraid). The *therapeutic goal* has two dimensions: the long-term goal of the entire therapy and the goal for a specific session. It is specified in the discussion between patient and therapist.

The information on all three aspects forms the basis for selecting appropriate music pieces and creating a specific therapy program for a patient.

Use Cases

1. Music Anamnesis and Preparation of a Therapy Program

Consider a patient with chronic pain. For him it is important to regularly use a music therapy program for pain relief. The patient answers questions in the music anamnesis through the music therapy application as preparation for the therapeutic session. The session is held as usual, but the therapist can begin with some knowledge of the patient's background since the anamnesis has been completed by the patient in advance. During the session, the therapist gets an impression of the mental state of the patient and they jointly define the therapeutic goal. The information in the anamnesis can be extended by the therapist in the session. This information is entered by the therapist into the system as part of general documentation which is stored in the electronic health record of the patient. After specifying the therapeutic goal, the system suggests music pieces supporting the therapeutic goal and considering the preferences of the patient as derived by the music anamnesis. The therapist can select music pieces that are compiled by the system later on to training programs for the patient at home. The system helps to select appropriate pieces and to create therapy programs.

2. Running the Therapy Program at Home

The same patient suffering from chronic pain starts the mobile application at home. The app asks the patient for his current mood and health status. Depending on the responses, the application compiles an appropriate program of suitable length based on pieces preselected by the therapist. The program is suggested to the patient. After listening to the music and some relaxation time, the system asks for feedback in terms of the mood of the patient or changes in symptoms after listening to the program. This information is forwarded to an electronic health record as a structured report that contains the date, length of listening, initial health state, and state after listening. In this way, an anti-pain ritual is created in the patient. The feedback information helps the system adapt its program creation strategy. The therapist also receives feedback on whether a music piece was helpful or not, which can be considered in the follow-up.

Requirements

From these use cases, we can derive the system objectives: to increase quality of life and provide music therapy for patients that can be continued at home. In terms of quality, the system will support patients in listening to an appropriate music program that is directed to the therapeutic goal. The economic value is that therapeutic sessions are accompanied by a continuous practice which can support the effects of therapy. The music anamnesis can be collected already in a digitized manner, available for the first therapy session.

We collected several functional and non-functional requirements. Functional requirements include:

- Supporting a music anamnesis
- Capturing current emotional or mental state or constitution of a person before and after listening
- Proposing a helpful music therapy program, taking into account recorded music history and current constitution of the person as well as the therapeutic goal and recommendation of the therapist
- Allowing listening to receptive music therapy program
- Collecting feedback and reconsidering emotional and mental state according to the therapy program (independent learning of the preferences of a person)

Non-functional requirements include: 1) standardized information modelling of the music anamnesis, 2) standardized reporting of home therapy sessions, 3) connection to an electronic health record to store the anamnesis and reports of therapeutic sessions at home, 4) consistent graphical user interface (GUI), also suited for patients less experienced with digital devices, and 5) response time less than 1 second.

Additionally, there are quality standards relevant for music applied in medical contexts [10] which need to be considered in the implementation of the system. This means, not each music piece is appropriate for usage in music therapy. Some relevant aspects are: a music piece needs to be defined (composing criteria, interpretation, instruments), selected according to the indication, should have a known profile of effect and the dosage needs to be specified (volume and duration of listening) [10].

Results

The system is designed as a service-oriented system architecture. Figure 1 shows the four major components of the architecture. It comprises a client-facing application acting as a user interface and a server application orchestrating single services. Processing services include the computational components, and resources encapsulate data that need to persist for the system to work properly. Details such as acquiring data from a client application or from multiple data sources of several types are not shown explicitly in the figure and will not be explained in detail. The components of the architecture are described below in more depth. First, a general overview is provided followed by a detailed description of single components.

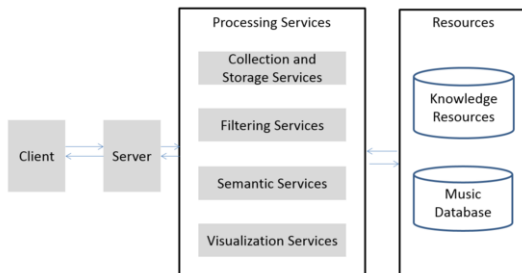


Figure 1 – System architecture comprising client, server, processing services, and resources

Architecture

The **Client** is a system user interface, partially designed as a conversational user interface to maintain a conversation-like style particularly in the anamnesis part. It allows starting the system, i.e. for patients this includes the anamnesis, and the therapy program. For physicians, the user interface presents the anamnesis, enables inputting the therapeutic goal or adapting the anamnesis, and supports compiling appropriate therapy programs.

The **Server** is responsible for interaction between client and processing services, invocation of services, and communication with external applications. In its role as service orchestrator, it invokes services in the right order and transmits the results of one service as input to the next service if required. It prepares responses of the processing services, collects results of the services and transmits the results to the client.

The **Processing Services** realise the actual processing. They might be independent from each other or the output from one processing service is required as input for another. We identified four different groups of processing services that are required to realise the music therapy app. These include: Data Collection and Storage Services, Filtering Services, Semantic Services, and Visualisation Services. Details are given in the following paragraphs.

Data Collection and Storage Services are responsible for collecting and storing data from and to different resources. One service retrieves the questions for doing the music anamnesis from a knowledge resource (see below) and forwards them to the GUI. Another service is needed for storing the completed music anamnesis into an electronic health record or for retrieving a stored anamnesis for a given patient. The data entries need to be stored in a structured manner for interpretation by the system later on in the music recommendation phase. Furthermore, it stores a report on the therapy program the patient received through the app including his feedback on health status in the electronic health record (EHR). It also stores information on the patient's mental state, frequency of listening, and music anamnesis – i.e. all information that is of relevance to the music therapist and which helps in monitoring progress. An additional service provided is retrieving music from music libraries or even data from a user's desktop as required.

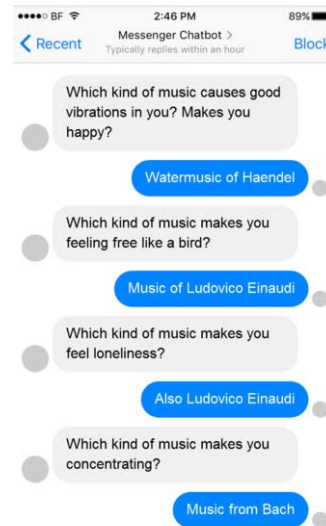


Figure 2 – Mock-up of the music anamnesis conversation

Filtering Services mainly aim at supporting the therapist in selecting appropriate music pieces from the database. Filtering criteria may include user-specified preferences as described in the music anamnesis, the specified therapeutic goal, etc. One potential filtering service is a recommendation service that makes suggestions for appropriate music pieces for a given patient or composes a therapeutic program suited for a specific patient given his/her mental and physical state. The therapeutic program that is created could for example be structured considering the phases in a therapy session where according to Zeitler [6] music of different characteristics are appropriate:

1. Salutation (music is rather slightly touching),
2. Warming-up (releasing music, more touching),

3. Confrontation (confronting music),
4. Closing and calm down.

For home usage, confronting music should not be included to avoid negative impacts.

Semantic Services provide music and text analysis functionalities on different levels of granularity. Their main task is to analyse music pieces, to select appropriate pieces when semantic information is unavailable. Additionally, semantic services are required for structuring and interpreting conversation with the patient through the conversational user interface. However, there are also services that are necessary to interpret the responses given by the patient in the conversational user interface.

The **Visualisation Services** realise the visualisation and result presentation of the application. Visual alerts, a conversational user interface, and scales for answering questions are some options that might be relevant. Additionally, the information from the music anamnesis is visualised to improve readability and perception of the information by the therapist.

Some of the processing services presented before require additional knowledge which is represented in the proposed

architecture as the **Resources** component. Two different kinds of resources can be distinguished. Knowledge resources contain background knowledge on a patient including the music anamnesis, therapeutic goal, and preferences in listening. This can be considered a patient information model. One knowledge resource is the questionnaire for music anamnesis. In order to recommend appropriate music to a given music anamnesis and therapeutic goal, corresponding knowledge needs to be available, i.e. knowledge on effects of music on patients in specific situations.

Music needs to be available to the system. It can be either integrated as resources or retrieved from external applications. For each music piece, knowledge on the effects it might cause and the type of music need to be available. Another option would be to develop algorithms that do such classification automatically.

Figure 3 shows the information flow for the use cases separated into three parts: music anamnesis, creating music therapy program, and music therapy. The structured report that is stored at the end contains information on the music pieces the patient was listening to, initial emotional and health state and feedback afterwards.

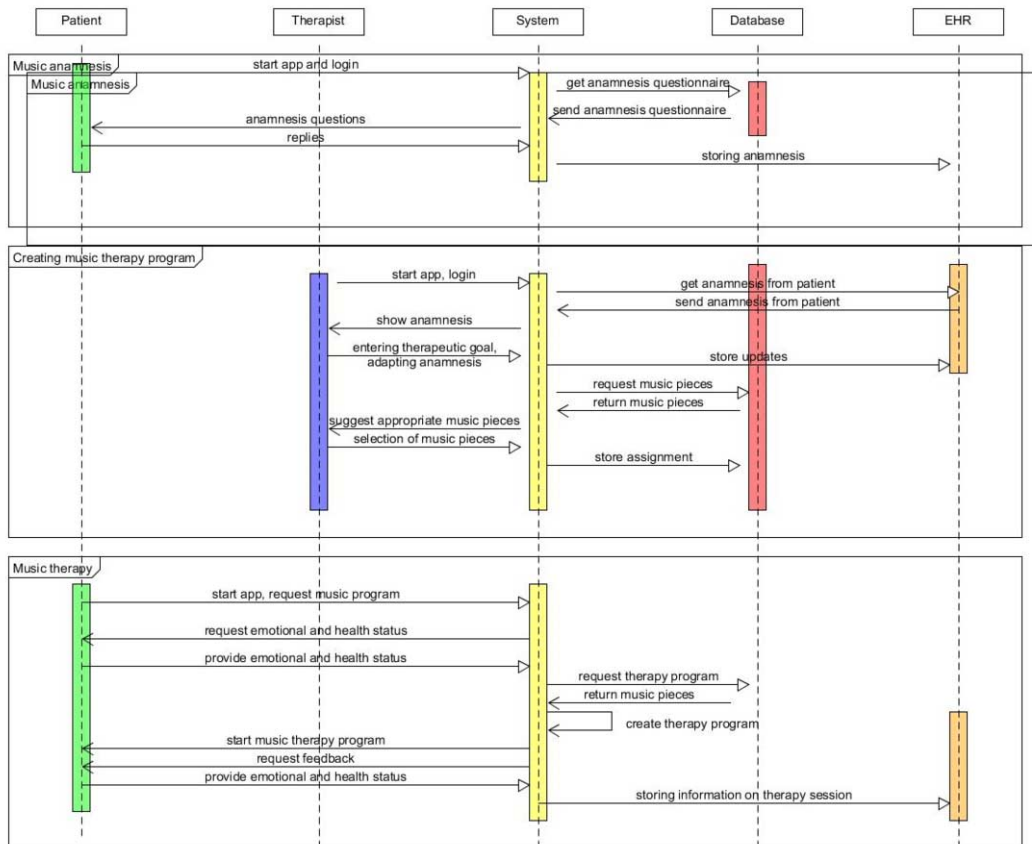


Figure 3 – Information flow diagram for music anamnesis, creation of music therapy program and consumption of music therapy

The music anamnesis will be realised using a conversational user interface, i.e. it resembles a short message function. This has the benefit of creating the impression of being interviewed by the therapist. Example queries in the music anamnesis screen is shown in Figure 2. Additionally, Figure 4 shows the user interface for starting the assessment of the current health state and mood.

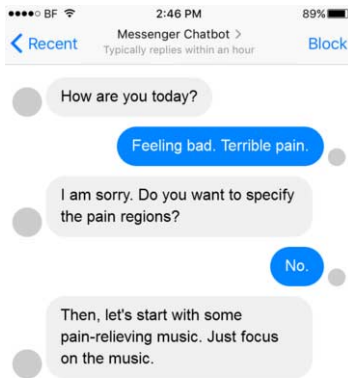


Figure 4 – Mock-up for starting the therapeutic music program

Discussion

In this paper, we introduced possibilities of information technology to support receptive music therapy. We introduced a concept and architecture for an IT-based music anamnesis and music therapy support. The next step is to evaluate the anamnesis tool with music therapists. It is clear that the app will not replace the therapist, but it could offer an opportunity for running therapy programs at home, like repeating exercises that have been introduced in a physiotherapy session. There are some open questions for future work. For the actual app, it needs to be clarified where the music is taken from. When using the music library of a patient, it would require an automatic classification of the music pieces in order to support the therapist and system in creating an appropriate program. Taking music from the internet also requires careful selection due to varying quality, interpretation, instrumentation etc. In order to automatically classify music, an open research question is how to formally describe the profile of music in terms of which effects or emotions are caused and whether it is possible to determine such categorization automatically. From a music analysis perspective, it is of interest to describe the relation between composition style and effect.

Secure data transmission needs to be ensured when transferring data to an electronic health record. Further, it needs to be checked whether this system should become a medical device.

The suggested approach tries to respect the current therapeutic procedure. The effect of music on the emotional and health state of a patient is difficult to predict and assess. The objective is to develop means to exploit the power of digital media to guide individuals to helpful, supporting music.

Conclusion

In this paper, we presented a concept and architecture of a mobile system that supports receptive music therapy. One aspect is the music anamnesis which can be filled in by a patient in a digital manner and can be extended during a therapy session. In the future, we will test the music anamnesis app and collect feedback from therapists for improvements. Beyond, the app is intended to enable home use of a music therapy program. It is still unknown to what extent home practicing supports receptive music therapy or improves quality of life for patients, such as those with chronic pain. The app will also allow the running of studies to assess this or similar questions.

References

- [1] L. Bunt, and B. Stige, Music therapy: an art beyond words. Routledge, New York, Second edition, 2014.
- [2] P. Vaiouli, K. Grimmer, and L.J. Ruich, Bill is now singing': Joint engagement and the emergence of social communication of three young children with autism, *Autism* **19** (1) (2015), 73–83.
- [3] K. Sherratt, A. Thornton, and C. Hatton, Music interventions for people with dementia: a review of the literature, *Aging & Mental Health* **8** (1) (2004), 3–12.
- [4] A.S. Maratos, C. Gold, X. Wang, and M.J. Crawford, Music therapy for depression, The Cochrane database of systematic reviews **1** (2008), CD004517.
- [5] X. Tan, C.J. Yowler, D.M. Super, and R.B. Fratianne, The efficacy of music therapy protocols for decreasing pain, anxiety, and muscle tension levels during burn dressing changes: a prospective randomized crossover trial, *J Burn Care Res* **31**(4) (2010), 590–597.
- [6] W. Zeitler, Auditive Music Therapy, tredition (Verlag), Hamburg, 2016.
- [7] R. Spintge, Musik in der Anästhesie und Schmerztherapie (Music in anaesthesia and pain therapy), *AINS* **35** (2000), 254–261.
- [8] M. Sommerer, Digital media in music therapy – an introduction. Musiktherapeutische Umschau online **34**(4), (2013), 351–357.
- [9] J. Nagler, Digital music technologies in music therapy practice. In: J. Nagler, Tomaino, C. M. (Hg.): Clinical Application of music in neurologic Rehabilitation, 2002, p. 41–49. USA/Saint Louis: MMB Music.
- [10] W.L. Magee, M. Bertolami, L. Kubicek, M. LaJoie, L. Martino, A. Sankowski, J. Townsend, A. Whitehead-Pleaux, and J. Zigo, Using music technology in music therapy with populations across the life span in medical and educational programs, *Music and Medicine* **3**(3) (2011), 146–153.
- [11] R. Spintge, Ausgewählte Aspekte und Grundlagen musikmedizinischer Anwendungen. (Selected aspects and basics of music therapy) In: R. Spintge (ed.): Musik im Gesundheitswesen. Neue Aspekte ganzheitlicher Medizin und Gesundheitsfürsorge, Schriftenreihe zur Gesundheitsanalyse **47** (2007), 7–26.

Address for correspondence

Prof. Dr. Kerstin Denecke, Bern University of Applied Sciences, Quellgasse 21, 2502 Biel / Switzerland; (kerstin.denecke@bfh.ch)

Use of Flowchart for Automation of Clinical Protocols in mHealth

Karine Nóra Dias^a, Daniel Welfer^b, Marcos Cordeiro d'Ornellas^b,
Carlos Jesus Pereira Haygert^c, Gustavo Nogara Dotto^c

^a Universidade Federal do Pampa, Alegrete, Rio Grande do Sul, Brazil

^b Department of Applied Computing (DCOM),

Federal University of Santa Maria (UFSM), Rio Grande do Sul, Brazil

^c Department of Clinical Medicine - Santa Maria University Hospital -
Federal University of Santa Maria (UFSM), Rio Grande do Sul, Brazil

Abstract

For healthcare professionals to use mobile applications we need someone who knows software development, provide them. In healthcare institutions, health professionals use clinical protocols to govern care, and sometimes these documents are computerized through mobile applications to assist them. This work aims to present a proposal of an application of flow as a way of describing clinical protocols for automatic generation of mobile applications to assist health professionals. The purpose of this research is to enable health professionals to develop applications from the description of their own clinical protocols. As a result, we developed a web system that automates clinical protocols for an Android platform, and we validated with two clinical protocols used in a Brazilian hospital. Preliminary results of the developed architecture demonstrate the feasibility of this study.

Keywords:

Clinical Protocols; Mobile Applications; Flowchart

Introduction

Software development brings together a number of basic macro-activities, which need specific techniques to execute them. Information technology (IT) professionals are responsible for performing these tasks in the software development process. The coding phase of a software is one of the main steps that requires a professional with technical knowledge of programming languages to consolidate the software as a product.

With the adoption of smartphones and tablets, the mobile applications have become popular to solve and to aid in problems of the personal daily and of diverse areas of the society. Consequently, one of the main sectors affected by the growing use of these information and communication technologies is health [1]. Therefore, physicians and other health professionals become users of mHealth and are not limited to the IT resources of health centers, which are precarious in many underdeveloped and emerging countries, as in several regions of Brazil.

While there are many mHealth applications available in the appStore, free of charge or paid, healthcare professionals rely on commercial or academic initiatives to develop health-specific applications once the construction requires specific skills for practitioners in the field. In addition, each existing application addresses a purpose that often does not serve a broad audience, as it is specific and / or limited to certain health care cases and institutions. In this way, the use of these tools

becomes unfeasible for a portion of health professionals, mainly to treat local problems from less open regions.

In the scientific literature are found several works that have been intended to develop mobile applications of Clinical Decision Support System [2], [3] and [5]. From these works, despite the divergent proposals, we can observe a pattern in the logic of development, which is in agreement with some representations of the clinical protocols.

The clinical protocol is a management tool that, in the form of systematized documentation or algorithms, normalizes the standard of health care at one point of attention [4]. The systematic reviews and assessments of the benefits and harms of alternative care options are the base to all content that constitutes the clinical protocols and therapeutic guidelines [5]. These practices aim at improving the quality of care, limiting unjustified variations in practice and reducing health costs, and transmitting health teams in the form of a document.

Given the importance of clinical protocols in the daily life of health professionals due to the contribution to care management based on structured and systematized knowledge, we propose to use them in the analysis phase to build applications. However, it is important to point out that each health institution elaborates its protocols according to factors and incidences of its locality, for example, the treatment of hospital-acquired pneumonia in a particular region of Brazil, will be different from the treatment proposed for this disease in China.

This work presents a proposal for automatic and generic generation of mobile applications, from the data entry structured in flowcharts. Remembering that the flowchart is an instrument used in the preparation of clinical protocols.

Background

The use of flowcharts for automation of clinical protocols is exploited as a form of computer-interpretable to provide decision-making systems. Classified as task-network model, these types of system formally represent clinical protocols, allowing an execution mechanism to process the knowledge represented in order to provide specific recommendations at point of care [6]. The main feature is the easy representation and human interpretation in graphic form. Many of these systems consolidated in the literature have the interpretation of the knowledge realized through own language for this purpose [7] and [8]. For the formulation of these languages are used primitive structures, logical and mathematical operators, flow control through if-then rules and ontologies. The use of each of these tools is dependent on their own syntax and semantics, in

which they work with groups of characteristics of the clinical protocols and temporal patterns, to control the patients' clinical situation. Some systems also include the organizational structure of the health system.

In this work, we use the perspective of these systems to carry out a knowledge-driven approach to specific clinical protocols applied to mobile technology. We emphasize that the structure of this work does not allow the temporal treatment of data, because the decision diagnoses are based on the information entered by the user, not being queried in databases, limited the configurations that will be presented in Methods.

Methods

The main feature of the proposal of this work is the form of user interaction with the system through flowcharts. In this way, the methodology used in this interaction, based on the definitions of the flowchart elements given by [6]:

- Oval element - Each algorithm begins with an oval design, representing a population of patients with a defined characteristic, symptoms and complaints. We can call these oval designs as clinical condition.
- Diamond - The diamonds represents the most important clinical decisions. They are decisive for the next steps and we call them of decision points.
- Rectangle - Represents the specific groups of the care process and the description of diagnostic or therapeutic interventions. We call them Service Process.
- Circle - Is the figure of closure, used as "output" every time a process reaches a conclusive stage. From this graphical element, we do not leave arrows.

We describe the proposed logic for these elements in automatic application generation as follows.

Clinical Condition

The clinical condition, represented by an oval symbol, is one of the simplest symbols. To describe it, only one clinical characteristic is necessary. Therefore, for the data entry in the system a text field is used. The clinical condition is obligatorily the first element of the flowchart, because from this the flow begins.

Clinical Decisions

Unlike the clinical condition, the clinical decisions, represented by the diamond, have several classifications in order to apply the most adequate logic to find the answer. Then, to configure the logic of the clinical decision element, the user must establish the most appropriate type of decision, at that point in the flowchart, in which it can be:

1. Numerical comparison.
2. Decision by calculation.
3. Existence of factors.
4. Question.
5. Scoring system.
6. Comparison from reference image.

Each option has a set of steps, which requires the necessary information for each decision logic. Below we describe the configurations of each type of decision.

Numerical comparison

To make a decision by means of a numerical comparison, we define as:

- Fixed values - where the comparison parameters will always be the same.
- Variable Values - where the comparison parameters vary according to the inputs.

It is important to emphasize the terms used in the proposed system and in the present work to determine the characteristics of input values, variables used in the formulation of the decision logic and description of the logic. We determine also:

- Comparison properties - The original terms in the clinical protocols. For the system, the comparison properties are essential information to compose the mobile application, because, through these terms, we identify the input fields of values. Example: Days of hospitalization, Temperature; Heart rate etc;
- definition of variables - such as fictitious terms to represent the comparison of properties in the formula of decision logic;
- Formula - the comparison structure used to carry out the decision logic. We standardize the formulas for each type of clinical decision chosen. Most often, they use the variables associated with logical operators. The user is responsible for describing the formulas according to the type of clinical decision setting chosen.

To compose the formula of clinical decisions based on numerical comparison of fixed or variable values, the logical operators are used.

1. In the description of the numerical formula of fixed values, the character x identifies it as a variable. Thus, for each type of numerical comparison with fixed values, a pattern is identified in the use of the x . For example, for comparisons with a fixed value the x will be at one end associated with a comparison operator and the fixed value, as in (1). For comparison between fixed numbers, x is in the center of the comparison, and the numbers in the borders, as exemplified in (2).

2.

$$x = 10(1)$$

$$0 \leq x \leq 20(2)$$

3. In another free text field, the user will describe in technical terms what represents the x , so that in the mobile application is described as input value, that is, the comparison properties.
4. The decision by numerical comparison with variable values is a situation where the outcome of the clinical decision is dependent on two or more variable values. Therefore, given the definition of two or more variables, we compare them as in the example (3):

$$y \leq x = z(3)$$

5. where y, x, z are variables which represent comparison properties.
6. Because of these clinical decisions, the logic will be binary, ie. two possible flows.

Existence of factors

Another way to make a clinical decision is to check for factors. In this way, the clinical protocols associate absence of factors and/or presence to decide the flow of the process. Including this characteristic in the system, this clinical decision we define through the textual description, in a single field, of these factors associated with logical operators, like "!" (NOT), "&&" (AND) and "||" (OR). We extract the factors from the logic of the exclusion of the operators and we use these words in the mobile

application. As a result of this type of clinical decision, logic will also allow only two flow possibilities.

Question

The clinical decision by question does not have any kind of computational logic. This decision depends on the interpretation of facts of the health professional, in which based on a questioning, will indicate if the flow will be positive or negative.

Decision by calculation

Like the numerical comparison of variable values, the decision by calculation will require the configuration of properties, definition of variables and description of the formula. However, we describe the formula as an equation through simple mathematical operators. We can use the result of the calculation described for conference (no flow option or only one) or establish classifications. To stipulate the classifications dynamically by the user through ranges of values, we use the logic of numerical comparisons with fixed value. The number of possibilities of flows will be (4). To define the flow for indeterminate ranges we add 1.

$$flows = ratings + 1(4)$$

Scoring system

The scoring system is an operation that takes into account a list of criteria, in which each criterion has a value. In order to establish the value of the criterion, we configure this value in fixed or the user must inform. For example, the criterion value is the same as the patient's age. In case the value is fixed, the treatment of the criterion is as the logic of the existence of factors.

By means of the sum of the values, of the selected criteria, we obtain the diagnostic result. To establish the result, in turn, we use the predetermined classifications. The number of possibilities of flows will be as in (4).

Comparison from reference image.

The comparison from the reference image does not require computational effort and is an intrinsic feature to other forms of clinical decision. Through the insertion of images, these can help the viewer user to make decisions.

Customer Service Process

The service process, represented by the rectangle, has the simple objective of providing information. It is usually associated with a clinical decision element. Therefore, it is the ideal element to describe textually the procedures given by the flow.

Simulator

The simulator has a structure, in the development interface, of logic and screens identical to those of the Android application. As the user mounts the flowchart, it is possible to check the expected outputs, according to the element types and their relationships. Therefore, the simulator is a way for the user to carry out case tests before the application is generated.

Results

To evaluate the efficiency and effectiveness of the proposed solution, we performed an experimental evaluation with

different clinical protocols of Brazilian hospitals. The evaluation explores divergent protocol perspectives applied in a mobile application. We implemented the logic of the flowchart elements in a web system for generating Android applications. We emulate the applications through Android Virtual Devices, version 2.2. The system default language is Brazilian Portuguese, but there is also a translation of the flowchart settings in English.

Protocol 1 - Treatment of community-acquired pneumonia (CAP)

Protocol 1 describes the treatment suggestions of the CAP through a scoring system, classifying them into 5 treatment possibilities: outpatient, with a lethality of 0.4%; Outpatient clinic, with 0.7% lethality; Hospital brief; Hospital, with lethality of 8.5%; and hospital with lethality 31.5%. In addition, in case of hospital treatments, we verify the need for intensive care unit (ICU). In all other cases, we verify the probability of specific germs and thus the appropriate prescription of drugs by doctors to fight the patient's illness happen. Subsequently, we realize the specific diagnosis of the pathogen. According to the documentation of protocol 1, in this situation, it is possible to transcribe in the proposed system all procedures, and mainly the decision logic. In this example, in the clinical decision element, the punctuation system and decisions by means of questions stand out.

Figure 1 illustrates the scoring system configuration where the user describes:

- The segment of characteristics - text that serves to structure and visually order the characteristics and their scores. A scoring system can have multiple segments, created dynamically by the left button "New Segment".
- Number of characteristics - To create dynamically rows in the table to describe the characteristic and its corresponding punctuation.

Characteristic	Score
Age in years - Woman	0
Age in years - Man	-10
Asylum residence	10

Figure 1 – Example of scoring system setup.

Later the creation of the configuration of the characteristics and their scores. The user must sort the result ranges. The number of result classifications indicated the limit of associations that the decision element might have. Figure 2 illustrates this process.

Classification	Score
Indeterminate (Mandatory classification if results do not fit within any range of values)	-
Ambulatory	=10
Hospital Breve	71-90
Hospital	91-130
Hospital	>130

Buttons: Add Classification, Save clinical decision data

Figure 2 – Example of setting the result ranges for the scoring system.

In the cases of clinical decision by question, the user is only responsible for verbatim informing the type of inquiry. The proposed system of the possibility of inserting an image to assist the physician in the inspection, as shown in Figure 3. However, this type of clinical decision can only be associated with two elements: one corresponding to the positive flow and the other to the negative flow.

Make clinical decision through question

Describe the question that should be displayed to the user:
Remember: Only the YES and NO options are displayed.
Risks for specif germs? See the table.

Assign a reference image to check the result. For example a table with a summary about the result obtained.
Add image: (Format accept - .jpg, .png, .gif, .psd, .bmp)
Escolher arquivo | table_risk.png | Upload

Patógeno	Fatores de risco	Tratamento
Pneumococo	pacientes idosos, com antecedente de fontes de risco específicas.	Penicilina G cristalina 4 a 8 milhões de UI por dia em ampolas 500
Pneumococo resistente à penicilina	uso de antibióticos β lactâmicos, quinolonas, colistina, ampicilina e cefalosporinas nos últimos 3 meses, doença imunossupressora, uso de corticóide a título suprativo a 65 mg ou em doses elevadas (até 360 mg), tratamento com fármacos nefrotóxicos para que penicilina deva ser a primeira linha de escolha para pneumonia pneumocócica	Penicilina G cristalina 2 milhões de UI de 6 a 8 horas em ampolas 1 g 500 mg (VTD)
Streptococcus pneumoniae	exatidão de idade, doença causadora de alteração da estrutura do pulmão pulmonar (DPOC).	Ampicilina, amoxicilina ou Acetaminofeno
Streptococcus pneumoniae produtor de β lactamase resistente a ampicilina	uso prévio de antibiótico β lactâmico nos últimos 3 meses;	Amoxicilina-clavulato
Enterobactérias	infecções recentes em instituições, doença cardíaca pulmonar, comorbidades médicas múltiplas, alcoolismo, uso recente de antibióticos.	Amoxicilina-clavulato
Pseudomonas aeruginosa	doença com comprometimento da estrutura do pulmão (ex: bronquiectasia), comorbidades médicas múltiplas (até 100 mg de prednisona por mês de 3 semanas), desnutrição, uso de antibiótico de amplo espectro.	Ofloxacina-clavulato
Aspergillus	doença aguda, doença neurológica, alteração de consciência, alcoolismo, distúrbio da digestão, aspiração alimentar durante o dia anterior.	Penicilina G cristalina 18 a 24 milhões UI por dia
Streptococcus aureus	disseminação, doença comprometimento da estrutura do pulmão (para MRSA: hospitalização recente).	Ofloxacina 2 g 4/6 horas, (Clonazepam se MRSA)
Legionella pneumophila	água contaminada	Acetaminofeno

Finish

Figure 3 – Example of configuration of the decision by question.

In the example presented, we structure the flowchart as shown in Figure 4. The transcription of all protocol 2 information in flowchart in the system, according to all the necessary configurations, took approximately 35 minutes. We use 1 clinical condition (Pneumonia acquired in the community); 4 elements of clinical decision, being a scoring system and three questions; 11 service processes; and a conclusion.

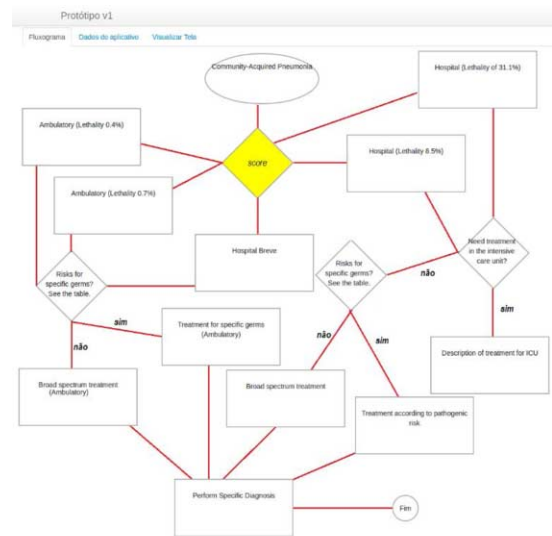


Figure 4 – Protocol 1 flowchart.

The application generated automatically through the flowchart has 16 screens, which reflect each element drawn. The flow of the screens is static except for the decision elements that require user interaction to point the next path. Figure 5 shows examples of each element used in the flowchart. Figure 5 [1] shows the characteristics of the scoring system. In 5 [2] a decision by question.

[1] Demographic Factor

Age in years - Woman:

Age in years - Man:

Sylum residence

Comorbidities

Neoplasm

Hepatopathy

Congestive heart failure

Cerebrovascular disease

Kidney disease

[2] Risk for specific germs? See the table.

Table with 4 columns: Fatores de risco, Descrição do fator de risco, Pontuação, and Tratamento.

Não Sim

Figure 5 – Examples of graphical interfaces generated for each type of flowchart element.

Protocol 2 - Management of infectious complications in severe acute pancreatitis

Unlike protocol 1 that explores a consolidated scoring system, we present the protocol 2 as a more explanatory document on the management of infectious complications in severe acute pancreatitis. In this way, the service process was widely used to structure the information in the application. Figure 6 illustrates the configuration of this element, in which it requires only: title, description and optionally image to elucidate some of the procedure.

Figure 6 – Example of configuration of the service process

For the diagnosis of pancreatic infection and systemic inflammatory response syndrome, we use the logic of factor checking. Therefore, for the clinical decision configured we chose existence of factors. In this type of decision, the user describes in text the factors to be verified and relates them through logical operators. It is also important to point out how many propositions indicated the positive diagnosis needs. Figure 7 illustrates, as an example of the configuration of this element, the diagnosis of pancreatic infection.

Figure 7 – Example of configuration of the decision through the existence of factors.

Finally, Figure 8 illustrates the graphical interfaces generated in the application from the respective elements exemplified above. In 8 [1] the graphical interface of a service process, and in 8 [2] the decision through the existence of factors are displayed.

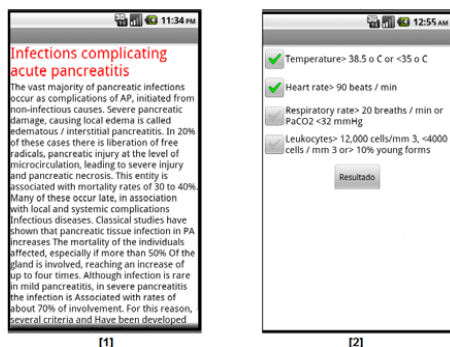


Figure 8 – Examples of protocol 2 graphical interfaces.

Conclusion

One of the ways to structure the protocol information is the algorithm. By means of flowcharts, it is possible to graph the logic of algorithms. In this work, we explored this characteristic of clinical protocols allied to the possible standard of many applications of clinical decision system. Thus, joining these aspects proposed a way to create generic mobile applications for health professionals, without the need for coding.

We obtained the experimental results through real clinical protocols. Due to the limitation of space, it is not possible to demonstrate all proposed features, however, for the examples presented it was possible to transcribe the information as a flowchart in an easy way. Due to the simplicity of the construction of flowcharts, the usability of the system becomes high, that is, there are no ambiguous interpretations for the elements or complicated technical configurations. In addition, the cost of development is quite low. Moreover, from the construction of the project, we realize the creation of an executable android file, without the user having to install any tools.

The generated android applications showed a simple graphical interface, but perfectly usable and correct logic according to the specification in the flowchart, which demonstrates the feasibility of the proposed structure. As future work, we aim to apply improvements in this project.

References

- [1] World Health Organization. K.S. Bonome. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. 2011.
- [2] D. Palazuelos et al. User perceptions of a mhealth medicine-dosing tool for community health workers. JMIR mHealth and uHealth, JMIR Publications Inc., Toronto, Canada, v. 1, n. 1, p. e2, 2013.
- [3] K. N. Dias, D. Welfer, J.F. Kazienko and R. Cassol. A novel ios m-health application to assist the hospital acquired pneumonia diagnosis and treatment. IEEE Latin America Transactions, IEEE, v. 14, n. 3, p. 1335–1342, 2016.
- [4] B. Martinez-Perez et al. Mobile clinical decision support systems and applications: a literature and commercial review. Journal of medical systems, Springer, v. 38, n. 1, p. 1–10, 2014.
- [5] M.A.F. Werneck, H.P. Faria, K.F.C. Campos. Protocolos de cuidado à saúde e de organização do serviço. Belo Horizonte: Coopmed, 2009.
- [6] R.M. Rosenfeld, R.N. Shiffman, and P. Robertson. Clinical practice guideline development manual, a quality-driven approach for translating evidence into action. Otolaryngology, Head and Neck Surgery, v. 148, n. 1 suppl, p. S1-S55, 2013.
- [7] M. Peleg. Computer-interpretable clinical guidelines: A methodological review. Journal of Biomedical Informatics, Springer, v. 46, n. 1, p. 744–763, 2013.
- [8] D.R. Sutton and J. Fox. The syntax and semantics of the PROforma guideline modeling language. Journal of the American Medical Informatics Association, v. 10, n. 5, p. 433–443, 2003.
- [9] M. Peleg, A.A. Boxwala, E. Bernstam, S. Tu, R.A. Greenes and E.H. Shortliffe. Sharable representation of clinical guidelines in GLIF: relationship to the Arden Syntax. Journal of biomedical informatics, v. 34, n. 3, p. 170–181, 2001.

Address for correspondence

K. Dias – karinenoradias@gmail.com
D.Welfer – daniel.welfer@ufsm.br

Tracking the Implementation of Electronic Medical Records in Dubai, United Arab Emirates, Using an Adoption Benchmarking Tool

Osama El-Hassan ^a, Amer Sharif ^b, Mohammad Al Redha ^a, Iain Blair ^c

^a Health Data and Information Analysis Department, Dubai Health Authority, PO Box: 4545 Dubai, UAE

^b Education Department, Dubai Healthcare City, PO Box 66566 Dubai, United Arab Emirates

^c Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, PO Box 17666, Al Ain, United Arab Emirates

Abstract

In the United Arab Emirates (UAE), health services have developed greatly in the past 40 years. To ensure they continue to meet the needs of the population, innovation and change are required including investment in a strong e-Health infrastructure with a single transferrable electronic patient record. In this paper, using the Emirate of Dubai as a case study, we report on the Middle East Electronic Medical Record Adoption Model (EMRAM). Between 2011-2016, the number of participating hospitals has increased from 23 to 33. Currently, while 20/33 of hospitals are at Stage 2 or less, 10/33 have reached Stage 5. Also Dubai's median EMRAM score in 2016 (2.5) was higher than the scores reported from Australia (2.2), New Zealand (2.3), Malaysia (0.06), the Philippines ((0.06) and Thailand (0.5). EMRAM has allowed the tracking of the progress being made by healthcare facilities in Dubai towards upgrading their information technology infrastructure and the introduction of electronic medical records.

Keywords:

Electronic Health Records, Health Services, Benchmarking

Introduction

A feature of health systems throughout the world over the past 20 years has been the increased use and reliance on Information and Communications Technology (ICT). Also called Health Information Technology (HIT) or eHealth, these developments are being pursued because of their potential to transform the delivery of health care by making it safer, more effective, and more efficient.

The application of ICT in health can be seen in various forms but one of the most recognizable is the Electronic Medical Record (EMR) or Electronic Health Record (EHR). Generally the EMR comprises a hardware and software platform that supports a dataset on each patient and which allows tests and treatments to be ordered and progress to be documented. More complex systems will support clinical decisionmaking and will allow information to be shared and reported. The EMR is a key part of a health information system (HIS) along with other subsystems such as human resources and finance [1].

Often the terms EMR and EHR are used interchangeably but there are important differences. The EMR is an organization-centric system that manages patient medical records from different sources that are used during a patient's care (diagnosis, treatment, lab tests, imaging, medication) while an EHR system is "inter-organization", connecting several EMRs

for the purpose of sharing information among the owners of those EMRs [2].

The goal of the EMR is a clinical environment that is paper-free. EMR systems are found at all levels from primary to continuing care and in all specialties. They vary in size and complexity from those serving a single doctor's office to systems involving groups of hospitals, clinics and other service providers. EMR systems have been extensively researched and a number of different research themes have emerged.

The installation of an EMR system is usually an expensive investment and early researchers focused on the net benefits in terms of cost saving and improved outcomes. In one tertiary hospital in a low-income country, EMR reduced length of stay, transcription times and laboratory costs resulting in a net benefit over five years of US\$ 613,681 [3]. In the United States (US), in an outpatient setting the introduction of EMR was estimated to have resulted in reductions in spending and increases in revenue of more than US\$ 952,000 compared with the year before implementation [4]. Also in the US, in a primary care setting, the net benefits from using EMR over a 5-year period was US\$ 86,400 for each provider due to savings in drug expenditures, better use of radiology tests, better capture of charges and decreased billing errors [5]. Implementing an EMR system at a general hospital in China produced a net benefit over a six-year period of US\$ 559,025 from a reduction in staff engaged in the creation of new patient records, fewer adverse drug events and dose errors, improved charge capture and decreased billing errors [6]. With respect to improved outcomes, a randomized clinical trial of an EMR-based clinical decision support system for diabetes showed that patients in the intervention group had significantly better diabetes control with increases in lifetime quality-adjusted life years (QALY) at a cost of US\$ 3,017 per QALY [7]. In a US trauma center, EMR increased attending surgeons' involvement in patient care, leading to increases in revenue and a reduction in hospital mortality [8]. Research has also focused on the implementation of EMR systems as this can be problematic. The successful implementation of an EMR system in a large Swedish teaching hospital was dependent on factors relating to the system of choice, clinician involvement, a realistic timetable and a willingness to change [9].

As the literature on the net benefits of EMR has accumulated, a number of systematic reviews have been published. The conclusions of these reviews have been mixed. One review found that while EMR may improve user and patient satisfaction, their impact on patient outcomes and quality of care was inconclusive [10]. A landmark report following an extensive review of the published literature on the benefits and

costs of EMR concluded that, while in individual pilot sites there was clear evidence that health care had been made safer, more effective and more efficient, the findings were not generalizable to other settings [11]. In particular there was a lack of knowledge about what specific systems and implementation methods should be adopted. Much of the positive evidence came from bespoke systems developed at academic medical centers or in very large health systems [12].

In many countries and in the US in particular the implementation of EMR is seen as a solution for the increasing costs of health care and as such has become a strategic goal [13]. In the US, the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act allows for incentive payments to increase adoption of EMR systems [14]. To qualify for these payments physicians must have “certified” or approved EMR systems and must be using them in a specially defined way, known as “meaningful use”, which includes at a minimum e-prescribing and electronic exchange of health information to improve quality. This requirement has led to considerable new research efforts to assess the proportion of providers that are achieving meaningful use and the characteristics of those that have done so [15]. Annual surveys are carried out and EMR adoption rates are published as performance indicators. Commercial and non-profit agencies have been established to which subscribers can report their progress on EMR implementation and thereby benchmark their performance against their peers and competitors. One such program is the EMR Adoption ModelSM (EMRAM) offered by HIMSS Analytics® [16]. The EMRAM is an eight-step process that allows providers to track their level of EMR adoption and make comparisons with other healthcare organizations. Originating in the US, EMRAM is now established in many other countries around the world.

The United Arab Emirates (UAE) is a federation of seven emirates and since its foundation in 1971 it has undergone rapid development and modernization. Mainly due to large net immigration of expatriates, over the last 10 years the population has more than doubled and in 2015 was estimated to be 9.2 million [17]. During the same period, the average life expectancy improved from 72 years to 77 years [18]. The UAE also performs very well on other international indicators for development and health [19].

Abu Dhabi, the capital of UAE, has its own health system while in the northern Emirates of Ajman, Fujairah, Ras Al Khaimah, Sharjah, and Umm Al-Quwain the Federal Ministry of Health retains a major role in the provision and regulation of health services in addition to its strategic UAE-wide role. The focus of this paper is on the Emirate of Dubai, which in 2015 had a population of 2.4 million and a modern infrastructure including excellent health services [20, 21].

Dubai Health and Medical Services (DOHMS) was established in 1970 as a local health authority and service provider. DOHMS continued to be the main local health authority in Dubai even after the formation of the federal Ministry of Health (MoH). In 2007, the Dubai Health Authority (DHA) was formed to oversee health strategy and regulation when it was separated from health service provision. Private health insurance is becoming the preferred funding source, although the Dubai government is still an important provider of services.

The DHA strategy is to develop efficient, high quality and accessible healthcare services for its population, and to achieve this it focuses on three critical areas: care delivery redesign, payment incentives and patient engagement. It does this in partnership with regulators, payers, providers and consumers, and the need to connect these health system players together

means that ICT and the management of medical information will play a key part in the delivery of the strategy [22, 23]. The potential of timely and accurate health information to improve population and individual health is clear [24]. However, the fragmentation of health information systems in the UAE is a challenge and leaders and decision-makers in the healthcare field often lack appreciation of the benefits of timely and accurate data [25]. Like many other countries Dubai has therefore developed an eHealth strategic plan to accommodate the changing role of ICT in improving health system performance and overcoming the current challenges. The Health Data and Information Analysis Department (HDIAD) is a directorate of DHA which focuses on both operational and strategic aspects of eHealth. In addition to providing decision makers in DHA and other government departments with health data for health system management, it also leads in the development and implementation of the strategic eHealth plan for Dubai. The purpose of this paper is to review how far healthcare organizations in Dubai have progressed in their automation journey from paper-based to paper-less health records management and the adoption of the Electronic Medical Record.

Methods

In this study we make use of data from the *Middle East EMR Adoption Model* (EMRAM), a self-reported survey tool that has been developed in a partnership between HDIAD and HIMSS Analytics, a global health IT organization [26]. Participants, comprising public and private hospitals in Dubai, submit their data by means of the EMRAM web-based tool and are then ranked at one of eight stages which cover the spectrum of EHR implementation from stage zero (a paper-based environment) to stage seven (an environment that is paper-free). Participants are also given an overall *EMRAM Score* which combines the current stage with credit for partially achieving higher stage applications. The requirements of each stage are shown in Table 1. HIMSS and HDIAD teams are responsible for setting up the EMRAM cycle or data collection round. They also follow up with participants to make sure that they complete the survey on time and to the required standards. The teams also conduct validation procedures to ensure the eligibility of participants who are ranked at stage five or above; these include site visits and telephone contact with chief information officers (CIO). Occasionally contact is required with software vendors involved in establishing or upgrading participants' EMR systems. Data collection rounds have taken place in 2011, 2012, 2013, and 2014, 2015 and 2016 (quarter 2).

Results

All public and private hospitals in Dubai are required to participate so the response rate is 100% for each data collection round. During the five-year period 2011-2016 the number of participating hospitals has increased from 23 to 33 (because new facilities have opened). In 2011, 21/23 (90%) hospitals were at Stage 2 or less and no hospitals were at Stage 5. Currently (2016, quarter 2), while 20/33 (60%) of hospitals are at Stage 2 or less, 10/33 (30%) have reached Stage 5 (Table 1). Between 2011 and 2016 the median EMRAM score increased from 1.2 to 2.6 and the proportion of hospitals achieving at least Stage 2 increased from 40% to 91%. This is important because Stage 2 is the minimum EMRAM stage required by Dubai Health Authority before a hospital can be considered for feasible integration with the Emirate-wide health data exchange platform. One hospital in Dubai has now achieved an EMRAM

Table 1. Stages of the EMR Adoption Model, number of Dubai hospitals achieving each stage, 2011-2016

Stage	Cumulative Capabilities	2011	2012	2013	2014	2015	2016
Stage 7	Complete Electronic Medical Record; Continuity of Care Document transactions to share data; Data warehousing; Data continuity with Emergency Department , ambulatory care and Out-patient departments	0	0	0	0	0	0
Stage 6	Physician documentation (structured templates), full Clinical Decision Support System (variance & compliance), closed loop medication administration	0	0	0	0	0	0
Stage 5	Full complement of Radiology-Picture Archiving and Communication System displaces all film-based images	0	2	6	6	10	10
Stage 4	Computer Practitioner Order Entry, Clinical Decision Support (clinical protocols)	1	2	1	1	1	1
Stage 3	Nursing / clinical documentation (flow sheets), Clinical Decision Support System (error checking), Picture Archiving and Communication System available outside Radiology	1	1	2	1	1	1
Stage 2	Clinical Data Repository, Controlled Medical Vocabulary, Clinical Decision Support , may have Document Imaging; health information exchange capable	7	10	10	12	12	17
Stage 1	Ancillaries - Lab, Rad, Pharmacy - All Installed	9	9	8	6	6	1
Stage 0	All Three Ancillaries Not Installed	5	3	1	3	2	2
	No. of participating hospitals	23	27	28	29	32	33

Source: HIMSS Analytics [23]

Score of 6.0 and has been nominated for Stage 6. Ratification of this status is awaited.

Discussion

The ultimate purpose of tracking EMR adoption among public and private sector healthcare providers is to effectively support the implementation of e-health capability. The results of Dubai EMRAM has shown that a substantial number of hospitals in the Emirate are making good progress with their automation journey. This could highlight the educational value of EMRAM in raising awareness of EMR implementation and fostering support for e-health capacity building.

The EMRAM staged methodology along with the survey results have allowed DHA to devise an evidence-based dissemination plan for the Health Information Exchange (HIE) platform. Based on the current EMRAM situation and the measured progress, DHA has set Stage 2 as the minimum requirement for healthcare facilities who wish to join HIE. Additionally, DHA has outlined a plan to leverage all facilities to stage 6 and to be able to effectively engage with the fully-fledged HIE platform which will be completed by the end of 2016. By that time, all patients in Dubai will enjoy continuity of care no matter which facility they visit.

It is notable that 30% of Dubai hospitals have reached Stage 5 but only one (having attained an EMRAM score of six) has been nominated for Stage 6. Stage 6 is among the most difficult to achieve because of the technical requirements for closed loop medication administration. Also, the stages above stage 5 have the highest impact on clinicians' workload and so require very careful implementation to ensure patient care is not disrupted. These are also the stages that have the greatest potential for return on investment, so for Dubai hospitals, as they contemplate the move to Stage 6, the future will be both challenging and rewarding.

The EMRAM surveys have the potential to improve the relationship between chief information officers (CIO) and chief executive officers (CEO). In our opinion there are three reasons

for this improvement: CIOs become able to clearly communicate the hospital automation situation with international and local benchmarks; both parties receive expert third-party feedback and assistance (from HIMSS Analytics) on how to improve their automation using tools such as gap analysis; and finally CEOs are able to exploit their hospitals' automation achievement in promoting services and winning market share.

In addition to tracking EMR implementation, EMRAM data has also allowed the benchmarking of Dubai hospitals against those in other countries with comparable health systems that participate in the HIMSS scheme. Dubai's median EMRAM score in 2016 (2.5) was higher than the scores reported from Australia (2.2), New Zealand (2.3), Malaysia (0.06), the Philippines (0.06) and Thailand (0.5). When using the "proportion of hospitals at each EMRAM stage" as the comparator, Dubai's performance is similar to that of Spain and Italy but exceeds Germany's. Compared to neighbor Saudi Arabia, the mean EMRAM score is similar but Dubai has a greater proportion of hospitals at Stage 5, albeit a smaller proportion at Stage 3.

There are limitations to the use of EMRAM. Firstly, EMRAM surveys do not report the expected improvements in achieved clinical outcomes within participating facilities which could follow the technological advances that are being made. Secondly, the self-reported approach is open to error unless validation procedures are carried out. These are resource intensive. Finally, from the eHealth planning point of view, questions seeking information about EMR integration capabilities within the survey are limited. To combat this, DHA uses a supplementary set of data collection to bridge this information gap.

Conclusion

Conducting EMRAM surveys within an active social network that comprises CIOs of Dubai's public and private hospitals has created a healthy environment for engagement between leading healthcare providers and the main healthcare regulator in the

Emirate, namely DHA. DHA has set standards for eHealth capacity, which hospitals are increasingly achieving. The goal is that all hospitals will have the capacity to join the Dubai Health Information Exchange so that health information will be shared between health providers to improve continuity of care, clinical outcomes and the patient experience.

Acknowledgements

The authors report no declarations of interest. None of the authors have received any specific support from any organization for the submitted work. No specific funding was required for this paper and ethics committee approval was not necessary as only secondary, aggregated data that is in the public domain is presented.

All the authors contributed to the ideas contained in this essay. OEH produced the first draft. All the authors refined the paper and edited the final version.

References

- [1] D Lewis, N Hodge, D Gamage, and M Whittaker, Understanding the role of technology in health information systems, *Pac Health Dialog* **18(1)** (2012), 144-54.
- [2] D Garets and M Davis, Electronic Medical Records vs. Electronic Health Records: Yes, There Is a Difference. Chicago, US: HIMSS Analytics; 2006 [cited 2201, Mar 23] Available from: http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSorg/Content/files/WP_EMR_EHR.pdf.
- [3] J Driessen, M Cioffi, N Alide, Z Landis-Lewis, G Gamadzi, O.J. Gadabu, and G Douglas, Modeling return on investment for an electronic medical record system in Lilongwe, Malawi, *J Am Med Inform Assoc* **20(4)** (2013), 743-8.
- [4] S. Barlow, J. Johnson, and J. Steck, The economic effect of implementing an EMR in an outpatient clinical setting, *J Health Inf Manag* **18(1)** (2004), 46-51.
- [5] S.J. Wang, B Middleton, L.A. Prosser, C.G. Bardon, C.D. Spurr, P.J. Carchidi, A.F. Kittler, R.C. Goldszer, D.G. Fairchild, A.J. Sussman, G.J. Kuperman, and D.W. Bates, A cost-benefit analysis of electronic medical records in primary care, *Am J Med* **114(5)** (2003), 397-403.
- [6] K. Li, S. Naganawa, K. Wang, P. Li, K. Kato, X. Li, J. Zhang, and K. Yamauchi, Study of the cost-benefit analysis of electronic medical record systems in general hospital in China, *J Med Syst* **36(5)** (2012), 3283-91.
- [7] T.P. Gilmer, P.J. O'Connor, J.M. Sperl-Hillen, W.A. Rush, P.E. Johnson, G.H. Amundson, S.E. Asche, and H.L. Ekstrom, Cost-effectiveness of an electronic medical record based clinical decision support system, *Health Serv Res* **47(6)** (2012), 2137-58.
- [8] D.L. Deckelbaum, A.J. Feinstein, C.I. Schulman, J.S. Augenstein, M.F. Murtha, A.S. Livingstone, and M.G. McKenney, Electronic medical records and mortality in trauma patients, *J Trauma* **67(3)** (2009), 634-6.
- [9] J. Øvretveit, T. Scott, T.G. Rundall, S.M. Shortell, and M. Brommels, Improving quality through effective implementation of information technology in healthcare, *Int J Qual Health Care* **19(5)** (2007), 259-66.
- [10] C. Delpierre, L. Cuzin, J. Fillaux, M. Alvarez, P. Massip, and T. Lang, A systematic review of computer-based patient record systems and quality of care: more randomized clinical trials or a broader approach?, *Int J Qual Health Care* **16(5)** (2004), 407-16.
- [11] P.G. Shekelle, S.C. Morton, and E.B. Keeler, Costs and benefits of health information technology, *Evid Rep Technol Assess (Full Rep)* **132** (2006), 1-71.
- [12] T. Garrido, B. Raymond, L. Jamieson, L. Liang L, and A. Wiesensthal, Making the business case for hospital information systems--a Kaiser Permanente investment decision, *J Health Care Finance* **31(2)** (2004), 16-25.
- [13] T.H. Payne, D.W. Bates, E.S. Berner, E.V. Bernstam, H.D. Covvey, M.E. Frisse, T. Graf, R.A. Greenes, E.P. Hoffer, and G. Kuperman, Healthcare information technology and economics, *J Am Med Inform Assoc* **20(2)** (2013), 212-7.
- [14] T. Burke, The Health Information Technology Provisions in the American Recovery and Reinvestment Act of 2009: Implications for Public Health Policy and Practice, *Pub. Health Rep* **125(1)** (2010), 141-145.
- [15] V. Patel, E. Jamoom, C.J. Hsiao, M.F. Furukawa, and M. Buntin, Variation in electronic health record adoption and readiness for meaningful use: 2008-2011, *J Gen Intern Med* **28(7)** (2013), 957-64.
- [16] HIMSS Analytics. Electronic Medical Record Adoption Model (EMRAM) [Internet]. Burlington, VT, USA: HIMSS Analytics; 2017. [cited 2017 Mar 22] Available at: <http://www.himssanalytics.org/emram>.
- [17] Department of Economic and Social Affairs Population Division. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. New York: United Nations; 2015. [cited 2016 Nov 15] Available at: http://esa.un.org/unpd/wpp/Publications/Files/Key_Findings_WPP_2015.pdf.
- [18] World Health Organization. World Health Statistics 2015. Geneva: World Health Organization; 2015 [cited 2016 Aug 11] Available at: http://apps.who.int/iris/bitstream/10665/170250/1/9789240694439_eng.pdf.
- [19] Human Development Reports United Arab Emirates. [Internet] New York; United Nations Development Program: 2016 [cited 2016 Oct 01] Available from: <http://hdr.undp.org/en/countries/profiles/ARE>.
- [20] E. Koornneef, P. Robben, and C. Hajat C et al. The development, implementation and evaluation of clinical practice guidelines in Gulf Cooperation Council (GCC) countries: a systematic review of literature, *J Eval Clin Pract* **21(6)** (2015), 1006-13. doi: 10.1111/jep.12337.
- [21] Estimated Population by Sector and Community [Internet]. Dubai; Dubai Statistics Center: 2016. [cited 2016 Nov 15] Available at: <https://www.dsc.gov.ae/Report/%20المنطقة%20المقدر%20حسب%20القطاع%202015.pdf>.
- [22] K. Lootah, J. Massey, and S. Amin, National Health Framework: The UAE Vision, *Proc of the 4th e-Health Conference in the Middle East* (2012), 167-172, [cited 2017 Mar 22] Available from: <http://www.innovationarabia.ae/wp-content/uploads/iarabiadoc/eHEALTH-2012.pdf>.
- [23] Dubai Annual Health Statistical Report 2015 [Internet] Dubai: Dubai Health Authority; 2016 [cited 2016 Nov 15] Available at: <https://www.dha.gov.ae/DHAOpenData/Annual%20Statistical%20Books/DHADoc768681140-28-07-2016.pdf>.
- [24] World Health Organization. Global Observatory for e-Health series, Volume 5: Legal frameworks for e-Health [Internet] Geneva: 2013 [cited 2016 Nov 15] Available at: http://www.who.int/goe/publications/ehealth_series_vol5/en/index.html.
- [25] E.J. Koornneef, P.B. Robben, M.B. Al Seiyari, and Z. Al Siksek, Health system reform in the Emirate of Abu Dhabi, United Arab Emirates, *Health Policy* **108(2-3)** (2012), 115-21.

[26] HIMSS Analytics Middle East [Internet] Singapore: 2013
[cited 2016 Nov 15] Available at:
<http://www.himssme.org/analytics/index.asp>.

Address for correspondence

Iain Blair (iain_blair@uaeu.ac.ae)

A Collaborative Evaluation Framework for Biometric Connected Devices in Healthcare

Troskah Farnia^a, Marie Christine Jaulent^a,
Guillaume Marchand^b, and Mobin Yasini^b

^aDepartment of Medical Informatics and Knowledge Engineering in e-Health, University Paris 6, Paris, France

^bDepartment of Research and Development, DMD Santé, Paris, France

Abstract

A large number of biometric connected devices are currently available with a variety of designs. Healthcare users cannot easily choose the reliable ones that correspond the best to their healthcare problems. The existing evaluation methods do not consider at the same time aspects of connectivity and healthcare usage. In this study, a collaborative evaluation framework for biometric connected devices in healthcare usage is proposed. This framework contains six dimensions: medical validity, technical reliability, usability, ergonomics, legal compliance, and accuracy of measurements. In a first step, these dimensions were assessed by designing a self-administered questionnaire answered by the stakeholders (patients, health professionals, payers, and manufacturers). A case study was then carried out in a second step to test this framework in a project of telemonitoring for heart failure patients. The results are in favor of the efficiency of the proposed framework as a decision making tool in healthcare usage.

Keywords:

Telemedicine; Decision Making; Delivery of Health Care

Introduction

The demand for wearable medical devices and remote patient monitoring systems is anticipated to increase over the years [1]. The wearable medical devices are a subset of connected devices. There is not a standard definition for connected devices [2]. The variation and the number of these connected devices are increasing to a predicted 50 billion devices by 2020, that is to say six times the human population [3]. Therefore, to decide and select the appropriate biometric connected devices that fit the healthcare user requirements and fully interoperate have become a critical issue. A comprehensive evaluation framework could help users to make better choices. To the best of our knowledge, there does not exist a comprehensive evaluation framework that evaluates and recommends specifically the biometric connected devices for various healthcare usages. There are just, some of these devices that have approved by CE medical mark. CE mark is a legal requirement to place a device on the market in European Union. CE mark does not present a holistic evaluation framework that covers all areas of assessment. It assesses particularly the quality, safety and reliability of the connected devices that may facilitate the ordering of these devices. However, there are different evaluation frameworks in telehealth and information communication technology (ICT) in healthcare such as: CHEF [4], GEMSA [5], MAST [6], KDS [7], CHEATS [8], TEMSED [9], 3Dimensions [10] and ASSIST [11]. They introduce other

assessment areas that could be complementary of CE mark assessment areas. Each evaluation framework provides a series of dimensions. Buccoliero [12] mentioned that the main ICT evaluation dimensions are stakeholders' point of views, costs, intangible corporate benefits and social benefits. However, the diverse telehealth evaluation frameworks cited above mention more issues: clinical, economical, technical, organizational and social as they are listed chronologically in the Table 1. However, the legal, educational and ergonomic aspects have not been much studied in these evaluation frameworks. These frameworks were applied in the various similar services including tele-consultation, tele-expertise, connected health ecosystem and elderly telehealth.

Table1-Telehealth Evaluation Dimensions

Evaluation Framework	Evaluation dimensions	Year
CHEATS	clinic, organization, administrative, education, social and technology	2002
3Dimension	quality, economic, education, clinic and accessibility	2005
MAST	clinic, technology, economic, legal aspects, patient perspectives	2010
TEMSSED	organization, economic, ergonomic technology, medical and social	2011
GEMSA	strategic, technology, service quality, economic, organization	2011
ASSIST	socio-economic	2012
KDS	economic, technology, socio-technical, policy, change management, ethical	2012
UAETA	accessibility, clinical, economical, technology	2013
CHEF	economical, change, socio-technical, end user perception, education, conformity	2016

In this study, the monitoring devices that collect and transfer the physiological data are considered as biometric connected devices. Inspired by these evaluation frameworks, the main objective of this study is to construct an evaluation framework in collaboration with service users for biometric connected devices in health usage.

Methods

In order to design an evaluation framework for connected devices addressed to healthcare usage, the main evaluation dimensions were selected according to telehealth evaluation frameworks in Table 1. This selection was established by an e-health expert group including two physicians (one of them also in evaluation of mobile health applications), one key opinion leader in telemedicine and health information systems, and one industrial engineer. This group defined the criteria for each dimension. Once the criteria had been developed, they were

integrated into a self administered questionnaire. There is one questionnaire for each evaluator profile (stakeholders: patient, health professional, service manager, etc.). Each questionnaire covers all dimensions for the targeted evaluator. This questionnaire was constructed in a collaborative way by various working sessions within the working group. Each criterion was then scored by each member of the working group according to its importance (from 1 to 5). The average score for each criterion determined the final importance index of each criterion. Some questions are addressed to users that include patients and healthcare professionals, other questions are addressed to service manager, payers (for the cases where users are not the payers) and device manufacturers that include managers and persons in charge of development and technical issues.

In order to quantify the qualitative answers of the questions, a score from 1 to 4 was assigned to each answer. Score “4” reflects the most favorable and satisfactory answer to the question, while score “1” is the most unfavorable response. For example, the question “Is this device easy to use ?” was assumed by a grid from 1 to 4, the score 4 is allocated if the device is very easy to use. On the contrary, the score 1 is chosen if the device is too hard to use.,

The final score was reported (we decided with the experts to set the score from 1 to 20) for each biometric connected device. It was calculated by the following equation (Eq. (1)):

(1)

$$\text{Connected biometric device score} = \frac{20 \times \sum(\text{importance index} \times \text{the average score given by the evaluators})}{\sum(\text{importance index} \times \text{maximum question score})}$$

This equation was used for five dimensions medical validity, technical reliability, usability, ergonomics and legal compliance. The sixth dimension “medical accuracy” was measured by the comparison of the error rate between the biometric connected devices and the existing validated professional medical devices. For example, the accuracy of the biometric connected “scale” was given by a number of measurements on the connected scale and at the same time on a validated and routinely used professional medical scale. The error rate was measured for each measurement and the average of these error rates were mentioned as the error rate of connected scale. For example, if the weight measured by the professional scale is 70.5 Kg and the average measures with the connected scale is 70.9 Kg, there is an absolute error of 0.4 and the percentage of error rate is 0.57 calculated by the following equation (Eq. (2)):

a: value measured by connected biometric device

b: value measured by professional device

$$\frac{(a - b) \times 100}{b} : \frac{(70.9 - 70.5) \times 100}{70.5} = 0.57 \quad (2)$$

This example concerns one user. The accuracy of the scale (percentage of error rates) is the average of at least 10 error rate percentages related to 10 users with various range of body weights.

To start a collaboration work and adapt the evaluation framework with the specific users’ context, a case study in a telemonitoring project for patients suffering from heart failure was carried out. This case study required another working group including two cardiologists, two nurses working in a cardiac service and one computer engineer. Various working sessions were organized to adapt the series of questions to heart failure patient use cases. 28 patients and 6 healthcare professionals participated to this case study by answering the

questionnaire. They have been trained to use the biometric connected devices and they have used them during one week. The connected devices included scale, blood pressure monitor, oximeter and pedometer. Two connected device manufacturers were selected for this test. For confidentiality reasons, the names of manufacturers were replaced by the symbols of X and Y. Other stakeholders of the project (project manager, device manufacturers) have also answered their dedicated questionnaires. The final score of each connected device was calculated from all responses of questionnaire.

To assess the accuracy aspect of biometric connected devices, we compared them with professional healthcare devices. Two nurses working in a cardiac service realized 45 measurements by each connected and professional devices. For each obtained measurement using two devices (connected and professional), the error rate was calculated. At the end of this process, the average of the error rates was shown as the error rate of the connected device.

There is no equivalent professional devices to measure physical activities. Therefore, the accuracy of pedometer has been measured by two physiotherapists riding on a treadmill with three speed levels.

Results

Six main evaluation dimensions were defined by the working group. For each dimension, various criteria were developed. Some examples of these criteria and the stakeholders to whom the criteria addressed are illustrated in Table 2.

Table 2 – Main Evaluation Criteria and Evaluator

Evaluation Dimensions	Criterion	Evaluator
Medical validity	concordance with medical guidelines cost, benefit, national policy, regional policy, measure units, relevance of sources,	Health professionals, Payer, manager
Technical reliability	technical characteristics, connectivity networks, maintenance, multi-users’ management	Device manufacturer, manager
Usability	motivation, satisfaction, acceptability, learning, knowledge	Patient, healthcare professional
Ergonomics	design, legibility, guidance, workload, adaptability	Patient, healthcare professional
Legal compliance	General terms and conditions of use, privacy and personal data protection respect	Manager, device manufacturer
Medical measurement accuracy	comparaison with healthcare professional devices	Healthcare professional

Evaluation dimensions are explained as follows:

Medical validity: It contains a series of questions that will be answered by the health professionals, payers and managers of service. It integrates many aspects: concordance with medical guidelines (including situations in which the devices are most useful, for example: wrist blood pressure monitors are not recommended unless the patient is overweight), cost/profit and national and regional policies.

Technical reliability: It explains the device characteristics, connectivity and maintenance requisites. The questions of this dimension will be answered by device manufacturers and service managers.

Usability: It evaluates the motivation, satisfaction, acceptability and technical knowledge level of users in using the connected devices. The users are the patients and healthcare

professionals, consequently they answer the related questions in this dimension.

Ergonomy: It evaluates the users' experiences in guidance (how the device guides the users for using it), legibility and workload (how the use of the device adds a workload to the users). The users are the evaluators of this dimension.

Legal compliance: It verifies the regulatory obligations and personal data privacy and security aspects. The most skilled persons who answered the related questions of this dimension are device manufacturers and service managers.

Medical measurement accuracy: Basically, the connected devices are manufactured for public and wellness use. The comparison between these connected devices and the professional validated medical devices that are used routinely by healthcare professionals allows to verify the accuracy of measures for connected devices in medical usage at patient's home.

In total, 49 questions were extracted from various criteria in all dimensions. Table 3 represents the number of questions in each dimension.

Table 3 – Number of Questions per Evaluation Dimension

Evaluation dimensions	Number of questions (%)
Medical validity	5 (10.2)
Technical reliability	14 (28.6)
Usability	14 (28.6)
Ergonomy	11 (22.4)
Legal compliance	4 (8.2)
Medical measurement accuracy	1 (2)

All of these questions are available under the request of the reader. Some examples of these questions are shown in Table 4.

Table 4 – Example of Evaluation Questions

Evaluation dimensions	Examples
Medical validity	Does the connected device (or the related documents) mention the sources of medical content and/or calculations provided ?
Technical reliability	Does the device have the ability to support multiple users?
Usability	Is this device easy to use ?
Ergonomy	The bluetooth/WiFi indicator that shows the device is connected, is it easily detectable?
Legal compliance	Is a legal notice provided with the device?
Medical measurement accuracy	What is the connected device error in comparison with professional devices?

Case study:

This evaluation framework was tested in the development phase of telemonitoring project for heart failure patients. Telemonitoring users (patients and healthcare professionals), manager and connected devices manufacturers answered the questionnaire for 4 connected devices. The final score of each connected device is presented in figure 1.

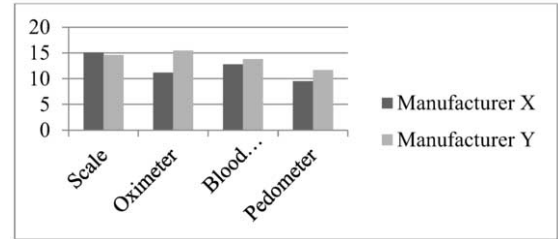


Figure 1- Final Score of Connected Devices

The final score of each connected device is obtained by the results of the questionnaire including five-evaluation dimensions: medical validity, technical reliability, usability, ergonomics and legal compliance.

Figure 2 shows the error rate percentage of three connected devices: scale, blood pressure monitor and oximeter.

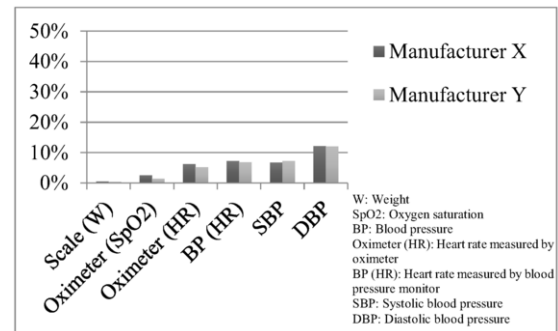


Figure 2- Error Rate of Three Connected Devices

The diastolic blood pressure error rate is very high (more than 10%), consequently this measurement will be less accurate. However, the weight measurements given by the two manufacturers' scales are very accurate (with almost 0% error rate).

Accuracy tests with pedometers were performed at three different speeds (3km/h, 4km/h and 5km/h). Figure 3 shows a very high error rate as 3km/h speed.

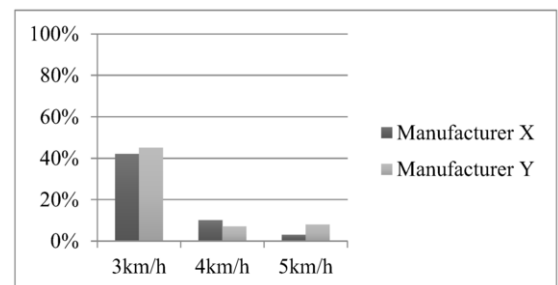


Figure 3- Error Rate of Pedometer in Different Speeds

The choice of pedometer depends on pedometer users' speed. For heart failure patients who walk maximum with 3km/h speed, both pedometers were not considered to be accurate.

According to these tests, the accuracy of the pedometers depends strongly on their wearing position. The pedometers were put in the pocket, on the belt, and on the wrist. Figure 4 shows that the pocket position has less error rate than the other positions.

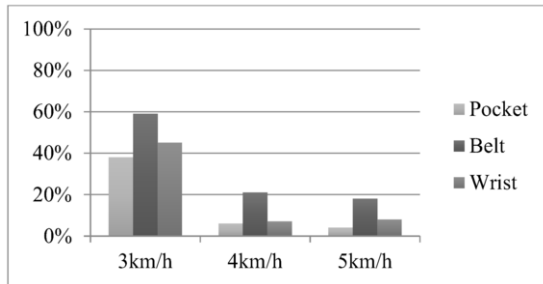


Figure 4- Error Rate of Pedometer in Different Positions

The final choice of the biometric connected health devices depends on their final score and their accuracy test. In this case study, the Scale of manufacturer X, the oximeter and blood pressure monitor of manufacturer Y obtained the most suitable score to the context of telemonitoring for heart failure patients.

Discussion

In this study, we proposed a collaborative evaluation framework for biometric connected devices in healthcare usage. This framework includes six dimensions that help users to make the best decision and choose the best devices among the large varieties that are adapted to their usage in a simple way. This framework can be easily adapted to different use cases. This process evaluation allows the manufacturers to better know their connected devices before embarking on the legal CE procedure.

In the literature we can find various studies on mobile health applications and mobile devices evaluation [13,14]. Our study specifically addressed the evaluation of the connected biometric devices and it involves all stakeholders in the evaluation process. It explains in details each criterion and the way to measure them by the scoring method. It also considers more specifically the accuracy of the measures that has not yet been mentioned in previous published studies.

The assessment of a connected device could not be carried out without the evaluation of the accompanied mobile application. This framework describes the dimensions and criteria regarding the connected device itself. An evaluation process for the accompanied mobile application should be added to this framework to assess the quality of the connected device in all aspects. An example of mHealth app assessment process that could be used beside this framework is mHealth Quality [13].

The final score showed that there is no great difference between the scale and blood pressure monitor of two manufacturers. However, the difference is considerable for oximeter and pedometer. Concerning pedometers, high error rates were seen in both analyzed devices. We have tested our framework only in two brands. This can limit the results of our analysis. However, these two brands were selected among the market leaders for connected biometric devices. Further research is needed to validate our results with a larger scale of connected devices brands and in the context of a broader chronic diseases.

This framework could be used in a broader multidimensional study for medical remote monitoring projects in chronic diseases. Another utility of this evaluation framework could provide the criteria to create a certification process for biometric connected health devices. The certification process needs to define some of the criteria as mandatory and set up the cut off score above which the devices could be certified. mHealth Quality [13] is actually working on this framework and use it in their certification process for connected biometric devices.

Conclusion

The use of wearable medical devices is certainly growing. Therefore, an individual is facing a jungle of connected items including very good devices as well as dangerous ones. A wide selection of mHealth devices with a real diversity of usages provides the opportunity to do almost everything in every medical specialty. However, this diversity prevents a "one size fits all" approach to ensuring the quality of these devices and information security. Exploring and creating adapted criteria to each device for providing individualized assessment method based on the provided functions is an important issue of the future research.

The results of this study show that there is still a long way to proceed to improve the reliability of the connected biometric devices particularly in pedometers. None of the devices reached the score 15 out of 20. To use these devices in the telemonitoring system, both health professionals and patients need to trust connected biometric devices. This trust can not be gained if these devices are not reliable in all discussed dimensions.

Acknowledgements

The Health regional agency of region Centre Val de Loire and France ANRT « Technology and research national association » provided financial support for this research. The authors thank Dr. Catherine Monpere team, Dr. Marc Goralski team, Dr. Paul Bardiere and Mr. Dominique Pierre.

References

- [1] Tillmann K. Understanding the connected health and wellness market. Consumer Electronics Association (CEA); 2014.
- [2] Moore Q, Johnson A. US Health care technologies. Baker Institute Publication; 2015.
- [3] Evans D. The Internet of Things : How the Next Evolution of the Internet Is Changing Everything. Cisco; 2011.
- [4] Carroll N, et al. Evaluating Multiple Perspectives of a Connected Health Ecosystem. Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies. 2016;5:17–27.
- [5] Le Goff - Pronost M, Picard R. Need for ICTs assessment in the health sector: A multidimensional framework. Communications and Strategies. 2011;83:87–108.
- [6] kidholm K, et al. A model for assessment of telemedicine applications: mast. international journal of technology assessment in health care. 2012;(28):44–51.
- [7] Khoja S, et al. Conceptual Framework for Development of Comprehensive e-Health Evaluation Tool. Telemedicine and e-health. 2013 Jan;19(1):48–53.
- [8] Brown M, et al. Evaluation Practices of a Major Canadian Telehealth Provider: Lessons and Future Directions for the Field. Telemedicine and e-health. 2008 Oct;14(8):767–74.
- [9] Fouquet Y, et al. TEMSED : une approche multicritère d'évaluation des technologies de télésanté et d'autonomie. Annals of Physical and Rehabilitation Medicine. 2011;(54):55–61.
- [10] Bashshur, R., et al. Telemedicine evaluation. Telemed E-Health. 2005;11(3):296–316.
- [11] Hammerschmidt R, Jones T. ASSIST -Assessment and Evaluation Tools for Telemedicine. empirica – Gesellschaft für Kommunikations und Technologieforschung mbH, Bonn, Germany; 2012.
- [12] Buccoliero L, et al. A methodological and operative framework for the evaluation of an e-health project. International

- journal of health planning and management. 2008;(23):3–20.
- [13] Yasini M, et al. mHealth Quality: A Process To Seal The Qualified Mobile Health Apps. *Stud Health Technol Inform.* 2016;205–9.
- [14] Villarreal V, et al. A Systematic Review for Mobile Monitoring Solutions in M-Health. *J Med Syst.* 2016;40(9):199–211.

Address for correspondence

Corresponding author: Troskah Farnia
Campus des Cordeliers
INSERM U1142 - LIMICS
15 rue de l'école de médecine
Esc D - 2ème étage
75006 Paris, France
Tel. (+33/0) 6 20 56 23 68 ; email farnia.troskah@gmail.com

Serious Games for Parkinson's Disease Fine Motor Skills Rehabilitation Using Natural Interfaces

Antônio Augusto Foletto^a, Marcos Cordeiro d'Ornellas^a, Ana Lucia Cervi Prado^b

^a Laboratório de Computação Aplicada, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil,

^b Hospital Universitário de Santa Maria, Universidade Federal de Santa Maria, Santa Maria, RS, Brazil

Abstract

Parkinson's Disease rehabilitation can be long and boring being difficult to maintain patient engagement on therapy programs. Novel technologies are allowing computer games to be played through natural interfaces. This paper presents the development and assessment of a system of serious games for fine motor skills rehabilitation using natural interfaces. The games were assessed through a questionnaire that evaluated the game experience through seven components: immersion, flow, competence, tension, challenge and positive and negative affect. In addition, a conceptual framework for development of serious games for fine motor skills rehabilitation was proposed. The results from the quantitative questionnaire suggested that the player experience was positive on all components assessed. Also, player experience between the three games was statistically the same, implying that the games can be used with consistency in a physical therapy rehabilitation program.

Keywords:

Computer Games; Motor Skills; Rehabilitation

Introduction

Parkinson's Disease (PD) is one of the most common neurodegenerative disorders, affecting round 1% of world population over 60 years old [1]. Among all the symptoms related to the disease, tremors, movement stiffness, lack of attention and postural instability can be highlighted. Tremors can be very severe, impairing the hands and fine motor skills of the subject. Fine motor skills represent muscle, bones and nerves coordination to execute accurate movements. These movements use hand and fingers to grab, manipulate objects and tools like when using a pencil or scissors. These fine skills can be impaired due to lesions, diseases, or even stroke; although spine, nerve, or muscle injuries might cause some disturbance as well. Subjects with PD can show difficulty to speak, eat or write because they lost some of their fine motor control [2].

Most treatments for PD are based on administration of drugs, but there is also recovery and rehabilitation by physical therapy programs. Physical therapy programs help to contain symptoms of the disease throughout physical exercises. These exercises can stimulate the brain allowing neurons to create new connections to fulfill difficulties that were imposed by the disease [3]. These therapy programs can take a long time, as the disease has no cure, and the exercises proposed can be boring and repetitive [4]. That can contribute to a drop on patient adherence to treatment or even evasion from the therapy program.

Technology can bring more to people's lives than simply improve performance on doing tasks, it can provide joy and richness, improving the experience. A Serious Game (SG), is a game that can transmit a message, knowledge, skill, or a content

to the player. In addition, it might improve the player experience through different kinds of interaction and on different contexts, e.g., health, training and education.

The main goal of serious games for health is to provide knowledge or enhance player skills as any other serious game but, in this case, it must serve a medical purpose. Serious games are being used in many health fields, from exergaming to surgery simulators. Videogames like Wii Fit and Kinect Sports have mainstreamed the genre among physical therapists. However, these games are not built for therapy and individuals need physiotherapy programs tailored to their needs. Also, this genre of games should have more flexibility to allow the therapy to be more personalized according to the subject's capacity and avoid further lesions [5].

Interaction in traditional games are mainly based on mouse and keyboard or joysticks. As computers decreased in size and cost and have growth in processing power, new platforms and interactions were developed. The possibility to create and interact with virtual worlds the same way we interact with physical objects has motivated many researchers to develop new technologies. In consequence, novel technologies that can track cerebral waves, eye movement, body movement, etc. are being developed. Movement sensors like the Microsoft Kinect or the Leap Motion allow the creation of natural interfaces. A natural interface is not an interface that is natural, but it makes the user perceive and use as natural. It is not a feature of a software or device to have a natural interface. The user must feel as natural while using that interface, through a natural interaction, using day-to-day gestures and movements.

Even when people with PD have access to a rehabilitation program with quality and qualified professionals, the program may be long and become unpleasant. In some cases, these individuals, may present lack of motivation during the program due to its degree of repeatability that causes tedium [4,6,7]. In addition, cases of depression and dementia on PD can withdraw these individuals from special care. The use of games has proved to be a good way to fight lack of motivation that results from repetitive exercises on therapy programs [8]. The works of [3, 7, 9, 10] are some that have explored development of serious games for health using movement sensors to help therapies for motor disorders. Although, there seems to be a lack of works focused on fine motor skills rehabilitation using hand movement sensors like Leap Motion. As an off the shelf device that can detect hand movements with low cost, Leap Motion should be explored as a tool for hand rehabilitation [6].

This paper describes the assessment of player experience in a system of serious games for health using natural interfaces. The games were tailored to help therapists on fine motor skills rehabilitation programs. A one-way ANOVA is used to analyze

if all the prototypes can provide the same game experience regarding *immersion, flow, competence, positive affect, negative affect, tension and challenge*. Additionally, based on the development of three prototypes, we propose a conceptual framework for development of serious games for motor skills rehabilitation using the Leap Motion device.

Leap Motion

The Leap Motion is a USB device that can detect hand and finger movements in a 3D environment with high precision and performance. The device is relatively small and robust, with a width of 80mm, height of 12.7mm and depth of 30mm (Figure 1). It has three infra-red emitters and two infra-red (IR) sensors inside. Its field of view (FOV) is round 61cm³ and it is shaped like a hemisphere. According to its website, the Leap Motion can track hand movements with an accuracy of 1/100mm, but [11] has found an accuracy of 0.7mm when tracking non-linear movements executed by a robotic hand. Nevertheless, it has higher accuracy than other devices on the market, e.g. Microsoft Kinect.

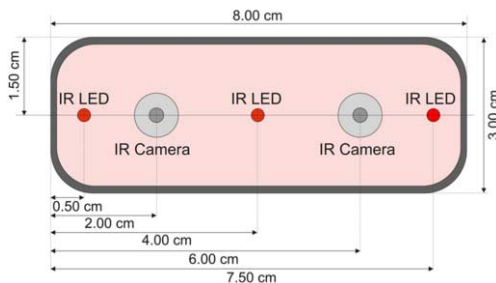


Figure 1 – A Representation of the Leap Motion Hardware Structure, Showing External Dimensions, Infra-Red Emitters and Sensor Positions [11].

Unity Game Engine

The Unity Game Engine was chosen mainly, because of its collection of tools and assets that allow rapid prototyping. Also, there is additional effort by Leap Motion developers to create bootstraps and tools for Unity. There is more documentation and also more graphic assets (prefabs) and scripts to use on prototypes. Moreover, Unity is the most used game engine by game developers according to data found on their public relations page.

Conceptual Framework

It is proposed a conceptual framework for the development of serious games with focus on motor skills rehabilitation using Leap Motion. This framework was based on a common architecture used to develop three prototypes. Basically, it has five layers: User layer, Input/Output layer, Game Engine layer, Database layer and Web Application layer. The framework proposed is illustrated in Figure 2.

The *User layer* is composed by the actors of the system: the player and the therapist. Both of the user profiles interact with the game through the I/O layer, but only the therapist should use the Application layer.

The I/O layer is responsible for the interaction/interface with the user. It is composed by hardware elements, like the Leap

Motion – responsible to provide hand and finger tracking data, Mouse and Keyboard, Sound and Display.

The Game Engine layer is responsible for all the game logic, since handling input/output to rendering the game. Usually, this can be done by a real game engine as Unity or Unreal. The components inside this layer are: *Leap Controller*, it processes the tracking raw data to be consumed by other components and objects; *Gesture Controller*, it checks tracking data for gestures and control gesture actions like grabbing or pinching; *Game Controller*, responsible for game management like scores, times, rules, etc.; *GUI Controller*, it manages the graphical interface elements of the game, gathers information from Game Controller and Database to output relevant information for the player and therapist; *Game Settings*, it manages the game adjustments that can be controlled by the therapist such as time counter, spawn speed, gesture sensitivity, number of objects to be spawn, etc.; *User Controller*, it is responsible for user data, such as user profile, user therapy program and sessions; *Database Controller*, handles the persistence layer of the game serializing information and storing in the database.

The *Database layer* is responsible for storing the data that is collected after each session. The Application layer can be represented by a Web Application that consumes data stored on the database to provide insights and report patient progress throughout the program.

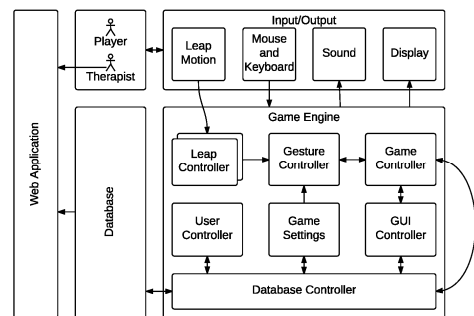


Figure 2 – Representation of a Framework for Development of Serious Games for Motor Skills Rehabilitation

This framework was created to cover the basic architecture for development of other serious games using leap motion. As three prototypes were developed using this conceptual approach, it can be argued that it can be used to develop more Leap Motion controlled serious games.

Prototypes

Three game prototypes were developed based on movements that were observed during rehabilitation sessions at the Rehabilitation unit of Hospital Universitário de Santa Maria (HUSM). In addition, physical therapy professors and professionals provided insights about fine motor skill exercises. These first prototypes were pitched to therapists, a game development professor and a design professor. Based on the considerations provided by those professionals, we iterated over the prototypes and refined them.

The prototypes were developed based on nature and farm life. This theme was chosen because the target audience is significantly composed by elders. The current generation of elders has at least some relation with nature and probably encountered a rural environment when young. Also, each game had a different metaphor that was chosen according to the gesture that was required to play in order to create affordance. The interaction has to make sense to the player, the gesture and the metaphor have to complete each other.

Game Prototype: Pinchicken

The mechanics of *Pinchicken* game is, basically, to pinch eggs that appear on the ground and drop it on the right chicken nest (Figure 3). The scene is composed by three chicken nests with chickens and eggs that keep falling on the ground. When the player pinch and egg, the game will highlight only one nest to drop the egg on. When the player succeeds moving the egg to the right nest, 10 points are awarded, a sound and a visual effect are played as feedback. When it is mistakenly placed, negative sound and visual feedback are played, but no points are awarded neither subtracted. Punishing the player is not the intention, because it can affect player motivation.



Figure 3 – Game 01: *Pinchicken*, Player Should use the Pinch Gesture to Grab Eggs.

Game Prototype: Finger-Hero

The prototype *Finger-Hero* is based on the mechanics of a blockbuster game called "Guitar-Hero". There are 4 lanes, each one has a flower at the end with a color (green, red, blue, yellow) and a thumb opposition gesture (index, middle, ring, pinky) associated. The game will randomly spawn bees that will move towards the flowers on each respective lane. The objective is to execute the correct opposition gesture when the bee is right above the flower. For example, if there is a blue bee on the third lane, as represented in Figure 4, the player should do the thumb opposition gesture with his ring finger when the bee is exactly over the blue flower. If he succeeds, 10 points are awarded and positive sound and visual feedback is given. Else if he misses only negative sound and visual feedback are given.

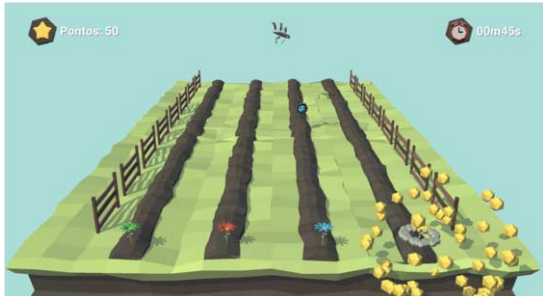


Figure 4 – Game 02: *Finger-Hero*, Player Should use Thumb Opposition Exercise to Play the Game.

Game Prototype: Grabduzeedo

On *Grabduzeedo* the player controls a spaceship with the hand and the tractor beam can be activated and deactivated using the grab and release gestures (Figure 5). The game will spawn sheep on a platform on the right side of the screen. The objective is to abduct the sheep, closing the hand (grabbing), and moving it to the platform with fences on the left. Just open the hand to release the sheep in place. When the sheep is placed inside the fence 10 points are awarded and positive sound and visual feedbacks are given. Otherwise, negative feedback is given, but no points are awarded or subtracted from the player.

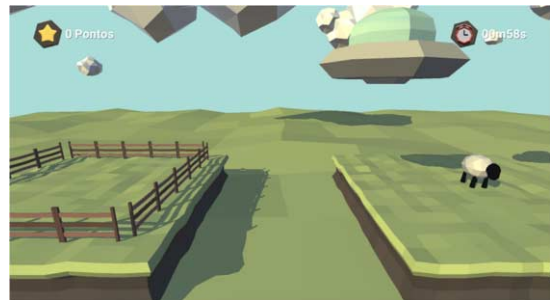


Figure 5 – Game 03: *Grabduzeedo*, Player Suse the Grab Gesture to Control the Spaceship's Tractor Beam.

Methods

Initial Setup

The experiments were conducted at the Rehabilitation Unit of HUSM, on a room with enough space and climate and light control. The last is extremely important because natural infrared light that is emitted from the sun can interfere on Leap Motion's tracking accuracy.

The setup includes a chair for the participant and desk to place the computer, the display and the Leap Motion device. The Leap motion was placed right in front of the display. Also, a camera was placed on a tripod in the corner of the room. The questionnaire and the consent form were printed in plain paper.

The game can finish when a countdown timer ends or when the player reaches a pre-defined maximum score. A time span of a minute and a half (1m30s) for each game was determined by the therapist as optimal to conduct the experiment. So, the participants had to sum the maximum points before the time ended. In addition, it was defined that the players would use only their dominant hand to play the games.

Game Experience Questionnaire (GEQ)

Game experience and common software experience are composed of different aspects and therefore they need to be assessed differently. The GEQ is an instrument to assess the overall game experience perceived by the player and it was proposed by [12]. They validated the questionnaire with 380 participants who played games of their own choice. The group was composed by 254 men and 120 women (6 null responses) who played games daily (29%), weekly (38%), monthly (13%), few times per year (12%) and hardly ever (8%). The instrument is composed of 4 modules: core module; in-game module; post-game module; and social module. The questions are answered using a Likert scale ranging from 0 to 4 (or from "not at all" to "extremely").

The core module has 33 questions and the result can be obtained by calculating the scores of seven components [12]: *Immersion*

is related to the fantasy, aesthetics, and imagination; *Flow* represents the experience of losing track of time and losing connection with the outside world; *Competence* express if the player was good, successful or skillful at the game; *Positive affect* represents if playing the game was fun or if it felt good; *Negative affect* is related to the player feeling bored, distracted or bad mood; *Tension* is related to how nervous, restless or annoyed the player felt; *Challenge* is related to how many effort the player think he put in the game or if he felt a time pressure.

The questionnaire and scoring system have some redundancies and spare items to prevent misunderstandings caused by translation. The in-game module has 14 questions (the questions are a mirror from the core module) and it is meant to be applied just after gameplay short breaks. The post-game module has 17 questions and assesses how the player felt after the whole gameplay session, its score is based on four components: *positive and negative experience*, *tiredness and returning to reality*. The social-presence module is meant for multiplayer games, aiming to understand how the player felt interacting with other player during gameplay through 17 questions.

For this study, only the core module was used. As the game sessions were short and single-player, the in-game, post-game and social-presence module were not required.

Participants

The participants were admitted for the experiment on a random and voluntary basis (with no monetary compensation) at the rehabilitation unit of HUSM. All the participants were from a physical therapy background: students, technicians, therapists and professors. Overall, we conducted the experiment with 20 healthy adults. According to [13] on a user experience or usability test only 5 people are sufficient to discover round 80% of the problems, but for quantitative purposes a number of 20 people is recommended. Also, [14] argues that most game evaluation works have a population of 11 to 20 people.

Procedure

First, all participants sat in a chair with the display and Leap Motion in front of them. They were instructed to sign the consent form needed for trials. Also, a brief explanation on how the experiment would proceed was given.

Before testing with the prototypes, the participants had a time to familiarize with the device and technology. A period to feel how the device tracks the hand, how far it reaches and how to interact with it. On this step, the participant interacted freely with the Leap Motion's built-in visualizer app.

After the participant was comfortable with the device, he proceeded to play the game. Exactly after each game had finished, the patient had to answer the questionnaire. This process was repeated for the other two games. The whole session took about 20 minutes, depending on each participant's speed to answer the questionnaire. In addition to the questionnaires, we recorded the scores of all the participants.

In order to analyze the data, GEQ-Scores were calculated. The score for each component is determined by the arithmetical average of the answers that are related to that component on GEQ's Manual. Then, the GEQ Scores of the games were compared with each other on every GEQ-Core component level. A one-way Analysis of Variance (ANOVA, $\alpha = 0.05$) was used to verify if GEQ scores had significant variances or if the game experience was the statistically the same across each game. The research hypothesis checked with ANOVA were:

- H_0 = The overall game experience for all games is the same;
- H_1 = At least one game has a different experience;

Results

The data from 37 questions was collected from 20 participants: 11 physical therapy students (55%), 1 professor (5%), 6 professionals (30%) and 2 technicians (10%); 16 Females (80%) and 4 Males (20%); Age ranging from 20 to 45 years old. Then, the score of each GEQ-Core component was calculated for each game: competence, immersion, flow, tension, challenge, negative and positive affect; and compared.

As it can be observed in Figure 6, optimal results were obtained for both positive and negative aspects of game experience for all the three games. Positive Affect, Immersion and Flow components had the best scores; Competence and Challenge presented a neutral/medium score while Tension; and Negative Affect had minimal score.

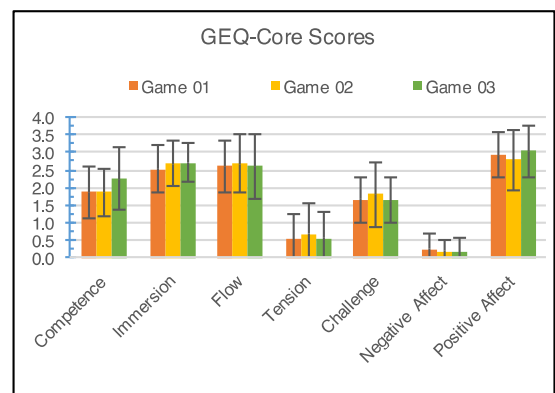


Figure 6 – GEQ-CORE Component Scores for each Game Prototype. The Y-Axis Represents the Scores and the X-Axis Represents each GEQ Component. Game 01 is Pinchicken, Game 02 is Finger-Hero and Game 03 is Grabduzeedo.

Player experience regarding Competence component resulted in statistically the same values for Game 01 ($M=1.86\pm0.73$) and Game 02 ($M=1.87\pm0.67$) and Game 03 ($M=2.25\pm0.92$). The ANOVA for all the components presented p-value superior to alpha, confirming the null hypothesis.

Discussion

It can be argued that the 'Competence' scores were positive because the games were somehow easy to play, allowing players to feel capable of completing the task. A slightly difference between values of Game 01 and Game 02 to Game 03 can be observed. That is because the gesture and interaction required to play the third game was much simpler than the others, requiring lower levels of coordination and dexterity to achieve the goal.

The best score values were from 'Positive Affect' component (Game 01 = 2.92 ± 0.65 ; Game 02 = 2.78 ± 0.85 ; Game 03 = 3.07 ± 0.74) it represents the fun and joy experience by the player during gameplay. Also, flow (Game 01 = 2.62 ± 0.75 ; Game 02 = 2.67 ± 0.85 ; Game 03 = 2.6 ± 0.92) and Immersion (Game 01 = 2.51 ± 0.70 ; Game 02 = 2.66 ± 0.66 ; Game 03 = 2.69 ± 0.56) presented good values representing that the game kept the player focused, was aesthetic pleasant and that it impressed the players. It is said that a game achieved immersion and flow when the player loses track of time and has the feeling to be transported to another place.

Overall, the results obtained from analyses of GEQ data suggest that the system of serious games prototypes presented has good level of user experience and usability on all three prototypes.

Further analysis with ANOVA pointed that the game experience across the three prototypes is statistically the same. That is a positive outcome because the prototypes were designed to work as a system of serious games, not as standalone games. Hence, it is expected that the game experience of each game is similar or the same. It provides consistency for the system.

This work is limited to player experience assessment. Further work should be conducted to evaluate the functionality and the validity of the games as a tool for rehabilitation. Also, there is still some space for improving on aesthetical and mechanics aspects of the game. Moreover, a web application that consumes data from database could be built to help the therapist track patient progress.

Conclusion

In this work, the main contributions are the development of a system of serious games for fine motor skills rehabilitation using NI. The system allows therapy sessions to be customized to patient needs, and also adding more fun and engagement. It can replace exercises that are repetitive and boring, helping to control patient evasion from therapy programs.

It is also proposed a conceptual framework for development of serious games for fine motor skills rehabilitation using Leap Motion. framework worked as expected during the development of three game prototypes.

Results obtained from the questionnaires point that the experience of the three prototypes are similar, allowing the therapist to use them on therapies with consistency. Positive components presented high scores and negative components scored near zero. Also these results can support a physical therapist to validate the system as a tool for rehabilitation of fine motor skills.

Acknowledgements

This research was supported by CAPES. The participation and support of all the individuals during the assessment experiments are thankfully acknowledged.

References

- [1] J.A. Muras, E.K. Stokes, and V. Cahill, Assistive technology in everyday living: a user survey of people with Parkinson's disease., *Technology & Disability* **20** (2008), 271-282.
- [2] N.K. Kaneshiro and D. Zieve, Fine Motor Control, in : *A.D.A.M. Medical Encyclopedia*, 2015.
- [3] D.J. Cargnin, M.C. d'Ornellas, and A.L.C. Prado, A Serious Game for Upper Limb Stroke Rehabilitation Using Biofeedback and Mirror-Neurons Based Training, *Stud Health Technol Inform* (2015), 348-352.
- [4] Y.-Y. Chao, Exergaming: therapeutic benefits in older adults, in: *Lower Extremity Review*, 2015, pp. 47-52.
- [5] S. McCallum, Gamification and Serious Game for Personalized Health, *Stud Health Technol Inform* (2012), 85-96.
- [6] J. Taylor and K. Curran, Using Leap Motion and Gamification to Facilitate and Encourage Rehabilitation for Hand Injuries: Leap Motion for Rehabilitation, in: *Advances in Medical Technologies and Clinical Practice*, IGI Global, 2015, pp. 78-104.
- [7] A.C. Macedo, R. Prada, and P.A. Santos, Serious Games for Motions Disorders Rehabilitation of Parkinson's Disease Patients, *Videojogos* (2015), pp. 10-16.
- [8] F. Laamarti, M. Eid, and A.E. Saddik, An Overview of Serious Games, *International Journal of Games Technology* (2014), 1-15.
- [9] R. McNaney, P. Olivier, M. Balaam, A. Holden, G. Schofield, D. Jackson, M. Webster, B. Galna, G. Barry, and L. Rochester, Designing for and with People with Parkinson's : A Focus on Exergaming, *Annual ACM Conference on Human Factors in Computing Systems* (2015), pp. 501-510.
- [10] I. Pachoulakis, N. Papadopoulos, and C. Spanaki, Parkinson's Disease Patient Rehabilitation Using Gaming Platforms: Lessons Learnt, *International journal of Biomedical Engineering and Science* **2** (2015), 1-12.
- [11] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, Analysis of the accuracy and robustness of the Leap Motion Controller, *Sensors* **13** (2013), 6380-6393.
- [12] K. Poels, Y.A.W. de Kort, and W.A. IJsselstein, D3.3: Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games, in: *Technische Universiteit Eindhoven* (2007).
- [13] J. Nielsen, How Many Test Users in a Usability Test?, in: *Nielsen Norman Group*, 2012.
- [14] A. Calderón and M. Ruiz, A systematic literature review on serious games evaluation: An application to software project management, *Computers & Education* **87** (2015), 396-422.

Address for correspondence

Antônio Augusto Foletto
afoletto@inf.ufsm.br
Av. Roraima, 1000, Prédio 7 – Anexo B, Sala 388
UFSM - Campus Universitário
97105-900, Santa Maria-RS – Brasil

Patient Portal Adoption Rates: A Systematic Literature Review and Meta-Analysis

Paolo Fraccaro^{a,b}, Markel Vigo^c, Panagiotis Balatsoukas^b,
Iain E. Buchan^{a,b}, Niels Peek^{a,b}, Sabine N. van der Veer^b

^aNIHR Greater Manchester Primary Care Patient Safety Translational Research Centre,
The University of Manchester, Greater Manchester Academic Health Science Centre, Manchester, UK.

^bMRC Health eResearch Centre, Farr Institute of Health Informatics Research, UK.

^cSchool of Computer Science, The University of Manchester, Manchester, UK

Abstract

Despite the increasing availability of online patient portals that provide access to electronic health records, little is known about their adoption by patients. We systematically reviewed the literature to investigate adoption of patient portals across studies. We searched MEDLINE and Scopus to identify relevant papers. We included 40 studies: 24 were controlled experiments, with prospective data collection in an actively recruited population; 16 were real-world experiments, with adoption being evaluated retrospectively after system deployment in clinical practice. Our meta-analysis showed an overall mean adoption rate of 52% (95% Confidence Interval [CI], 42 to 62%). Rates differed markedly between study types: controlled experiments yielded a mean adoption rate of 71% (95% CI 64 to 79%), compared to 23% (95% CI, 13 to 33%) in real-world experiments. This difference was confirmed in a meta-regression analysis of the influence of study characteristics on adoption rates. Our findings suggest that adoption rates reported in controlled studies do not reflect those in everyday clinical practice. Until we understand how to effectively increase adoption, patient portals are unlikely to consistently lead to improvements in care processes and health outcomes.

Keywords:

Computers/utilization; Personal Health Record; Patient Access to Records

Introduction

Patient portals are online services that allow people to have access to their electronic health records (EHRs) and support basic activities such as booking appointments, recording symptoms, and communicating with healthcare providers [1, 2]. They are considered a valuable instrument to engage patients in having a more active role in their care [3–8] and aid self-management [8–10]. Many patient portals are developed for people with long-term conditions [9, 11].

Patient portals are increasingly available [1, 12, 13], but their impact on health outcomes has yet to be established [14–18]. Previous systematic reviews found positive effects on patient engagement and satisfaction [14, 15, 17, 18], but evidence on the effect of patient portal use on care processes and health outcomes is conflicting [14, 15, 18]. Our understanding of this variation in impact is currently limited [14, 15]. This might be related to the fact that published evaluations tend to focus on

clinical endpoints, without considering the complex process that leads to them [19, 20].

As proposed by Coiera in his ‘information value chain’ [20], for any health information system to have impact, users first need to adopt the system and effectively interact with it (step 1) in order to receive information (step 2), which might then influence their decision making (step 3). This might lead to improved care processes (step 4), and, under beneficial conditions, to better health outcomes (step 5). Although not sufficient, good results are necessary at each step to achieve eventual positive impact on clinical endpoints [20]. Following Coiera’s information value chain, evaluating adoption rates is essential for understanding the (lack of) effect of patient portals on decision making, care processes and health outcomes.

Although individual patient portal studies may have reported on adoption rates [21–23], up till now no study has summarized these rates across studies. Also, adoption rates in controlled experiments may not always translate well to a real life context [15]. This disconnect between findings from controlled and real-world experiments might partly be explained by the former being more likely to include strategies for optimizing recruitment and minimizing attrition than the latter [24]. For example, a study in primary care showed that patients in practices with a more active strategy to promote portal use were more likely to be registered users compared to those in practices with a less active or no strategy [25].

Therefore, we systematically reviewed the literature on patient portals to investigate adoption rates across individual studies, and how rates might differ between controlled and real-world experiments.

Methods

We followed the PRISMA statement [26] to design and report our systematic review and meta-analysis, where applicable.

Search strategy

We used the search strategy that we developed for a broader literature review on methodological approaches to evaluate patient portal usage, usability and effect on decision making (i.e. Coiera’s information value chain steps 1-3). We searched for English language articles in MEDLINE via Ovid and in Scopus by combining subheadings and text words for patient portals with those referring to system usage, usability, and decision-making (full searches available at [27]). The searches were performed on the 18th of July 2016.

Selecting relevant studies

We considered studies relevant if they:

- Evaluated patient portals, using the definition of Irizarry *et al.* [2]: systems providing patients with access to their EHRs, and allowing them to enter health data or share information with their healthcare providers. We excluded studies on systems only providing educational material, or online booking or secure messaging functionalities.
- Had patients, carers, or healthy volunteers from the general population as the study sample, excluding those testing the system only within the research team.
- Reported sufficient information to determine portal adoption rate in the study population. Adoption was defined as the % of eligible patients who logged in at least once or—if this information was not available—had an active account during the study period. We excluded studies that only reported other types of usage statistics (e.g. frequency of use).
- Collected data on adoption through system interaction logs. Studies retrospectively asking patients about their portal use in surveys were excluded because such data lack in objectivity [28] and are known to be affected by recall bias [19]. We also excluded studies investigating intended use.
- Were peer-reviewed original articles (including conference papers) or systematic reviews in English, while excluding conference abstracts, narrative reviews, editorials, view point papers and grey literature.

After removing the duplicates from the Ovid and Scopus searches, the principal reviewer (PF) independently screened the titles and abstracts of all studies, whereas two secondary reviewers (PB; SvdV) did 50% each. For studies considered potentially relevant, we retrieved the full papers and two reviewers independently identified those meeting our inclusion criteria. At each stage of the review process, disagreement was solved through discussion.

Data extraction

We built a data abstraction form on the basis of previous systematic reviews on patient portals [17, 29] and health information systems evaluations [19, 30, 31], and pilot-tested it among the authors (PF, SvdV, MV, NP). The final form included items related to: general study characteristics; study type (controlled versus real-world experiment); study population; patient portal functionalities (data access; data recording; data sharing); number of potential patient portals users; number of patients who logged in at least once or had an active account during the study period. For real-world experiments, we additionally extracted information on strategies aimed at increasing patient portal adoption.

Controlled experiments were defined as any study where patients were actively recruited to participate in research and where data were collected prospectively. This included experiments with single groups as well as those with multiple groups (with participants being allocated either randomly or non-randomly), with the number of potential users equalling the number of participants recruited into the study. In case of multiple groups, we calculated this number across groups. Real-world experiments concerned observational studies that retrospectively evaluated adoption after a patient portal had been deployed in clinical practice. In this case, the number of potential users were all eligible patients in the clinical context of interest who should have been offered access to the portal.

One author (PF) extracted the data and performed the data synthesis for all studies. Uncertainties during the data extraction process were addressed and resolved by discussion with a second member of the research team (SvdV).

Data synthesis and analysis

We organised results according to study type. We calculated the mean adoption rate and 95% confidence interval (CI), both overall and stratified for study type. We also conducted a random-effects meta-regression analysis to evaluate the influence of the study and patient portal characteristics on adoption rates using the *metafor* R package [32]. Prior to the regression analysis, we performed a logit transformation on adoption rates reported in the included studies [33]. From the meta-regression coefficients estimates, we calculated Odds Ratios (ORs) and 95% CI. Furthermore, as prescribed by Stevens *et al.* [33], coefficients estimates were back-transformed to the proportional scale by using the mean adoption rate across all studies as anchor value. This gave us the estimated effect of each model covariate in a study with an average adoption rate.

Results

Figure 1 shows the flow diagram of our screening and selection process that yielded 40 relevant studies.

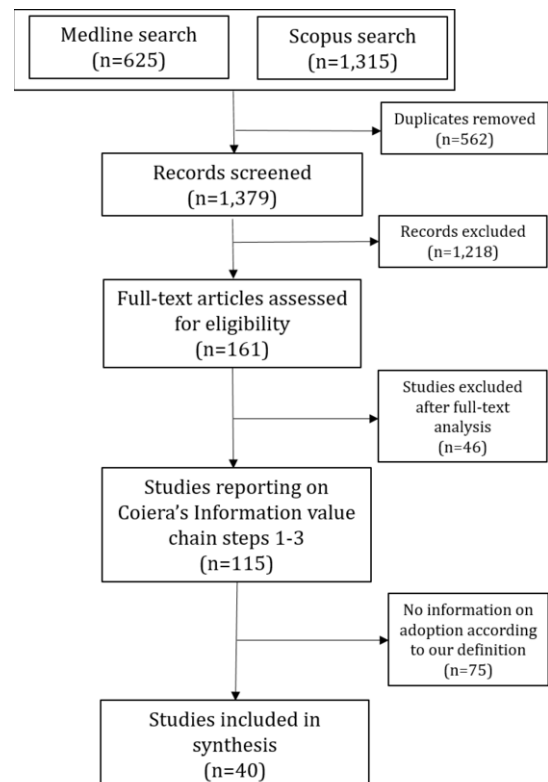


Figure 1 - Flow diagram of selecting relevant studies.

Study characteristics

Table 1 presents the general characteristics of included studies (extracted information and the full reference for each individual study is available at [27]). We identified 24 controlled and sixteen real-world experiments. The majority of included studies were published after 2010, conducted in the United States, and had a duration of 1 to 2 years. The vast majority of patient portals under evaluation provided some kind of access to patients' EHR as well as to data recording functionalities. Overall, the two study types had similar characteristics, with the main difference being the number of potential portal users. In particular, we found smaller numbers of potential users in controlled compared to real-world experiments, with the majority reporting values below 1,000 and above 10,000 people, respectively.

Table 1 - Characteristics of included studies and patient portal functionalities (N=40).

	Controlled experiments (N=24)	Real-world experiments (N=16)
<i>Year of publication</i>		
< 2005	2 (8)	/
2005 to 2010	5 (21)	4 (25)
> 2010	17 (71)	12 (75)
<i>Geographical location</i>		
United States	17 (71)	11 (69)
Europe	4 (17)	4 (25)
Other	3 (13)	1 (6)
<i>Study population</i>		
Patients with a specific condition	7 (29)	6 (38)
General population	6 (25)	3 (19)
Primary care	6 (25)	6 (38)
Other	5 (21)	1 (6)
<i>Number of potential users^{a)}</i>		
<100	8 (33)	/
100 to 1,000	13 (54)	1 (6)
1,000 to 10,000	2 (8)	5 (31)
> 10,000	1 (4)	10 (63)
<i>Study length</i>		
<1 year	8 (33)	2 (13)
1 to 2 years	13 (54)	9 (56)
> 2 years	2 (8)	5 (31)
<i>Main patient portal functionalities^{b)}</i>		
Data access	20 (83)	15 (94)
Data recording	18 (75)	12 (75)
Data sharing	10 (42)	6 (38)

^{a)} Calculated as: number of patients recruited into the study for controlled experiments; number of people eligible to be offered access to the patient portal for real-world experiments.

^{b)} Categories are not mutually exclusive.

Strategies to increase adoption in real-world experiments

Five of sixteen (31%) real-world experiments did not report any strategy to promote adoption. The remaining eleven reported a range of strategies aimed at increasing patient portal adoption rates. In five studies, eligible patients were directly invited (i.e. via mail or staff) or provided with system credentials that they only had to activate. Four studies disseminated promotional material to raise people's awareness of the portal via different channels, such as flyers and posters in waiting areas, directed electronic mailings or via staff. In three studies, computers were available in clinical settings for patients to access the portal. Lastly, two offered user training, one had personnel available onsite to support patients with using the portal, and one offered a help desk service.

Patient portal adoption rates

The overall mean adoption rate across all included studies was 52% (95% CI, 42 to 62%). When stratifying for study type, controlled experiments had a mean adoption rate of 71% (95% CI, 64 to 79%), compared to 23% (95% CI, 13 to 33%) for real-world experiments.

The results from the meta-regression are shown in Table 2. Study type was the only statistically significant covariate in the model after adjusting for all other study and patient portal characteristics. In particular, the OR for controlled versus real-world experiments was 10.8 (95% CI 3.2 to 36.3). This would translate to a difference in adoption rate of 40% (95% CI, 26 to 46%).

Table 2 - Estimates of ORs for adoption and change in adoption rate (%). Bold coefficients are statistically significant.

Study characteristics and portal functionalities	ORs [95% CI]	Change in adoption (%) ^{a)} [95% CI]
Study length (per year increase)	0.7 [0.4,1.4]	-7 [-21,8]
<i>Study type</i>		
Real-world experiment (reference category)		
Controlled experiment	10.8 [3.2,36.3]	40 [26,46]
<i>Publication year</i>		
<2005 (reference category)		
2005 to 2010	0.4 [0.0,7.0]	-24 [-50,36]
>2010	1.0 [0.1,18.7]	0 [-46,43]
<i>Geographical location</i>		
Others (reference category)		
United States	1.3 [0.3,5.9]	7 [-27,34]
<i>Study population</i>		
General population (reference category)		
Patients with a specific condition	0.7 [0.1,4.2]	-7 [-39,30]
Primary care patients	0.7 [0.1,3.6]	-9 [-39,28]
Other	1.4 [0.2,9.5]	8 [-34, 39]
<i>Patient portals functionalities^{b)}</i>		
Data access provided	0.5 [0.1,3.3]	-17 [-45,26]
Data recording provided	0.6 [0.1,2.2]	-14 [-38,19]
Data sharing provided	1.9 [0.5,7.9]	16 [-18,38]

Abbreviations: CI, Confidence interval; ORs, Odds ratios.

^{a)} Change in adoption for a study with underlying adoption rate equal to 52% (i.e. mean value across all included studies).

^{b)} Each of the three functionalities was included in the model as a binary variable, with 'Data access/recording/sharing not provided' as the reference category.

Discussion

Summary of the findings

We performed a systematic review of the literature and a meta-analysis of adoption rates of patient portals. The overall mean adoption rate was 52% (95% CI, 42 to 62%). Despite the majority of real-world experiments reporting the use of strategy to promote adoption, rates evaluated in real life were markedly lower compared to those evaluated in controlled experiments (23 and 71%, respectively); most of the other study characteristics and patient portal functionalities were similar between study types. A meta-regression analysis confirmed these findings.

Relation to other studies

Our meta-analysis is the first comprehensive, systematic review of studies reporting patient portal adoption rates. We are also the first to provide a summary statistic of adoption rates across individual studies, while quantifying the influence of study type and other characteristics on the rates reported.

Our findings complement what was found by Irizarry *et al.* [2]. Whereas they reported on the barriers to adoption of patient portals (i.e. involving personal characteristics, system usability, provider endorsement and security), our study provides evidence on the size of the problem of suboptimal adoption.

Our study confirms what was suggested by Giardina *et al.* [15] in their systematic review of randomised controlled trials (RCTs) of patient portals, who hypothesised that evaluation of adoption in clinical practice may give different results from those obtained in RCTs. Two main reasons may underlie their hypothesis. First, patients who agree to participate in this type of study are generally interested in and motivated to use the patient portal [34]. Second, patients who are recruited for a research study know that their actions will be studied, which is likely to change participants' behaviours (i.e. Hawthorne effect). Both reasons may increase the levels of adoption.

What is the meaning of the findings and what are their implications?

In the studies that we labelled real-world experiments, it was more difficult to assess the denominator than in controlled experiments. For the former, the denominator was commonly set as the largest group of patients who could have used the portal, while for the latter it was the number of patients to whom the portal was actively offered as part of the study. Due to this difference in counting, real-world experiments were more likely to have lower adoption rates, which may have resulted in our overestimating the difference between the two study types.

As prescribed by Coiera's information value chain [20], to increase the probability of improving care processes and health outcomes we first need to ensure that the previous steps in the chain have been taken successfully. Our study showed that when deployed in a real-world context, most patient portals failed to obtain high adoption rates. It is noteworthy that our findings were based on a rather crude definition of adoption (i.e. at least one login or activated account during the study period). However, in order for patients to receive substantial information from portals that might affect their decisions, more may be required than just activating a user account or logging in only once. Furthermore, applying a more sophisticated definition of adoption (e.g. logging in multiple times over a sustained period of time) is expected to have resulted in even lower adoption rates for both study types.

Low adoption rates were obtained despite real-world experiments reporting to have used active strategies to maximise enrolment and facilitate portal use. Therefore, the problem might not only be related to if and how patient portals are promoted, but also to whether patients consider portals relevant for self-managing their condition in everyday life. In this regard, a positive example comes from Kaiser Permanente, which was the only real-world experiment obtaining high adoption rates (i.e. 62%) in a large population [35]. Kaiser Permanente is the most widely used privately owned patient portal in the world [36]. They have their patient portal at the centre of their business model. Patients, in addition to gaining access to their EHRs, can use the platform

to download documents and forms necessary to use Kaiser Permanente services [35]. A previous study further showed how patient-centeredness and making information actionable are other main components of their success [37].

Limitations

This systematic review has two main limitations. First, some steps of the review process were not performed by two independent researchers. Therefore, these steps might have been more prone to errors than others; it was up to the discretion of the primary reviewer to discuss items that were less straightforward to extract and required more interpretation. Second, with 40 studies included, the meta-regression analysis might have been underpowered to detect statistically significant effects for factors other than the study type. However, as shown in Table 1, study characteristics were similar between the two types of studies; therefore we do not expect this to have influenced our results.

Conclusion

Overall, studies on patient portals found that half of the targeted population adopted the intervention. However, this number was markedly lower when evaluated in real-world experiments, with only one in four patients adopting the portal once deployed in clinical practice. Therefore, patient portals are unlikely to influence clinical endpoints in a real-world setting. Future studies in this field should focus on identifying factors and processes that positively affect adoption of patient portals in clinical practice.

Acknowledgements

Funded by the National Institute for Health Research Greater Manchester Primary Care Patient Safety Translational Research Centre (NIHR GM PSTRC) and the MRC Health eResearch Centre, Farr Institute, UK (MR/K006665/1). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- [1] C. L. Goldzweig et al., Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review., *Annals of internal medicine* 159(10) (2013): 677–87.
- [2] T. Irizarry, A. DeVito Dabbs, and R. C. Curran, Patient Portals and Patient Engagement: A State of the Science Review, *J Med Internet Res* 17(6) (2015): e148.
- [3] C. Pearce and M. Bainbridge, A personally controlled electronic health record for Australia., *Journal of the American Medical Informatics Association : JAMIA* 21(4) (2014): 707–13.
- [4] Institute of Medicine (USA), Achieve Meaningful use - Stage 2. <http://www.healthit.gov/providers-professionals/step-5-achieve-meaningful-use-stage-2>.
- [5] S. Urowitz et al., Is Canada ready for patient accessible electronic health records? A national scan., *BMC medical informatics and decision making* 8(1) (2008): 33.
- [6] C. Pagliari, T. Shand, and B. Fisher, Embedding online patient record access in UK primary care: a survey of stakeholder experiences, *JRSM Short Reports* 3(5) (2012): 34.
- [7] B. Fisher, Patients' access to their electronic record: offer patients access as soon as you can., *The British journal of general practice : the journal of the Royal College of General Practitioners* 63(611) (2013): e423-5.
- [8] J. Adler-Milstein et al., A Comparison Of How Four Countries Use Health IT To Support Care For People With Chronic Conditions., *Health affairs (Project Hope)* 33(9) (2014): 1559–66.
- [9] C. S. Kruse et al., Patient and Provider Attitudes Toward the Use of Patient Portals for the Management of Chronic Disease: A Systematic Review, *Journal of Medical Internet Research* 17(2) (2015): e40.
- [10] M. Price et al., Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review, *BMC Medical*

- Informatics and Decision Making* 15 (2015): 32.
- [11] M. Price et al., Conditions potentially sensitive to a Personal Health Record (PHR) intervention, a systematic review, *BMC Medical Informatics and Decision Making* 15(1) (2015): 1–12.
 - [12] D. Wiljer et al., The anxious wait: assessing the impact of patient accessible EHRs for breast cancer patients., *BMC medical informatics and decision making* 10(1) (2010): 46.
 - [13] C. Bartlett, K. Simpson, and A. N. Turner, Patient access to complex chronic disease records on the Internet., *BMC medical informatics and decision making* 12 (2012): 87.
 - [14] E. Ammenwerth, P. Schnell-Inderst, and A. Hoerbst, The impact of electronic patient portals on patient care: a systematic review of controlled trials., *Journal of medical Internet research* 14(6) (2012): e162.
 - [15] T. Davis Giardina et al., Patient access to medical records and healthcare outcomes: a systematic review., *Journal of the American Medical Informatics Association : JAMIA* 21(4) (2014): 737–41.
 - [16] S. de Lusignan et al., Patients' online access to their electronic health records and linked online services: a systematic interpretative review., *BMJ open* 4(9) (2014): e006021.
 - [17] F. Mold et al., Patients' online access to their electronic health records and linked online services: a systematic review in primary care, *British Journal of General Practice* 65(632) (2015): e141–e151.
 - [18] C. S. Kruse, K. Bolton, and G. Freriks, The effect of patient portals on quality outcomes and its implications to meaningful use: A systematic review, *Journal of Medical Internet Research* 17(2) (2015).
 - [19] A. W. Kushniruk and V. L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, *Journal of Biomedical Informatics* 37(1) (2004): 56–76.
 - [20] Enrico Coiera, *Guide to Health Informatics* (CRC Press, 2015).
 - [21] S. G. Smith et al., Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort, *Journal of the American Medical Informatics Association : JAMIA* 22(4) (2015): 888–895.
 - [22] D. Masys et al., Giving patients access to their medical records via the internet: The PCASSO experience, *Journal of the American Medical Informatics Association* 9(2) (2002): 181–191.
 - [23] G. R. Phelps et al., Patients' Continuing Use of an Online Health Record: A Quantitative Evaluation of 14,000 Patient Years of Access Data, *J Med Internet Res* 16(10) (2014): e241.
 - [24] A. Zweben, L. M. Fucito, and S. S. O'Malley, Effective Strategies for Maintaining Research Participation in Clinical Trials, *Drug information journal* 43(4) (2009): 10.1177/009286150904300411.
 - [25] C. K. Yamin et al., The digital divide in adoption and use of a personal health record, *Archives of Internal Medicine* 171(6) (2011): 568–574.
 - [26] A. Liberati et al., The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration., *PLoS medicine* 6(7) (2009): e1000100.
 - [27] Paolo Fraccaro et al., Supplementary material for "Patient portal adoption rates: a systematic literature review and meta-analysis." <http://doi.org/10.5281/zenodo.439927>.
 - [28] J. A. Bargas-Avila and K. Hornbæk, Old Wine in New Bottles or Novel Challenges: A Critical Analysis of Empirical Studies of User Experience, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011): 2689–2698.
 - [29] C. Y. Osborn et al., Patient Web Portals to Improve Diabetes Outcomes: A Systematic Review, *Current diabetes reports* 10(6) (2010): 422–435.
 - [30] K. Hornbæk, Current practice in measuring usability: Challenges to usability studies and research, *International Journal of Human-Computer Studies* 64(2) (2006): 79–102.
 - [31] P.-Y. Yen and S. Bakken, Review of health information technology usability study methodologies, *Journal of the American Medical Informatics Association : JAMIA* 19(3) (2012): 413–422.
 - [32] W. Viechtbauer, Conducting Meta-Analyses in R with the metafor Package, *Journal of Statistical Software: Vol 1, Issue 3 (2010)* (2010).
 - [33] S. Stevens et al., Analysing indicators of performance, satisfaction, or safety using empirical logit transformation, *BMJ* 352 (2016).
 - [34] J. A. Casey et al., Using Electronic Health Records for Population Health Research: A Review of Methods and Applications, *Annual Review of Public Health* 37(1) (2016): 61–81.
 - [35] N. P. Gordon and M. C. Hornbrook, Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: A database and survey study of seniors in a large health plan, *Journal of Medical Internet Research* 18(3) (2016).
 - [36] L. W. Johnson et al., Successful Practices in the Use of Secure E-mail, *Permanente Journal* 18(3) (2014): 50–54.
 - [37] T. Otte-Trojel et al., The organizational dynamics enabling patient portal impacts upon organizational performance and patient health: a qualitative study of Kaiser Permanente, *BMC Health Services Research* 15(1) (2015): 1–12.

Address for correspondence

Corresponding author: Paolo Fraccaro, MEng.

Email: paolo.fraccaro@manchester.ac.uk.

E-HOSPITAL – A Digital Workbench for Hospital Operations and Services Planning Using Information Technology and Algebraic Languages

Daniel Gartner^a and Rema Padman^b

^a School of Mathematics, Cardiff University, Cardiff, Wales, United Kingdom,

^b The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Abstract

In this paper, we describe the development of a unified framework and a digital workbench for the strategic, tactical and operational hospital management plan driven by information technology and analytics. The workbench can be used not only by multiple stakeholders in the healthcare delivery setting, but also for pedagogical purposes on topics such as healthcare analytics, services management, and information systems. This tool combines the three classical hierarchical decision-making levels in one integrated environment. At each level, several decision problems can be chosen. Extensions of mathematical models from the literature are presented and incorporated into the digital platform. In a case study using real-world data, we demonstrate how we used the workbench to inform strategic capacity planning decisions in a multi-hospital, multi-stakeholder setting in the United Kingdom.

Keywords:

Models, Theoretical; Decision Making; Hospital Administration

Introduction

The rapidly growing patient population worldwide and the increasing demand for high-quality healthcare services are imposing severe capital, resource and human capacity constraints on hospitals. For example, one in every five Medicare beneficiaries in the United States is hospitalized at least once or multiple times per year. On the supply-side, almost 5,000 inpatient, acute-care hospitals exist nationwide that treat these beneficiaries. Of the approximately \$300 billion dollars spent on the Medicare program per year, almost \$100 billion is spent on inpatient services [1].

Given limited budgets, hospitals seek to treat patients efficiently and effectively in order to stay profitable. Adapting inpatient services that aim to improve the planning of hospital-wide workflows using information technology (IT), operations management (OM), and advanced data analytics (DA) techniques are some of the recent developments that we observe in healthcare delivery [13, 15, 16, 28, 29].

In this paper, we demonstrate this convergence by proposing a unified digital workbench to help multiple stakeholders to improve the planning and allocation of scarce hospital resources to improve transparency and efficiency of inpatient services. Additionally, we demonstrate the feasibility of the proposed workbench by applying it to capacity planning decisions at a multi-hospital site using a preliminary prototype implementation.

Hierarchical Modelling of Organizational Decision Making

We draw on the classical hierarchical management decision levels [2] to delineate different stakeholders' objectives for a using our workbench at each decision-making level. A framework to break down business decisions into strategic, tactical and operational decision levels are illustrated in Figure 1 [2]. Its essential aim was to assess the environment of an organization and to adjust internal resources accordingly [11]. The model is depicted by the regular triangle shown in Figure 1.

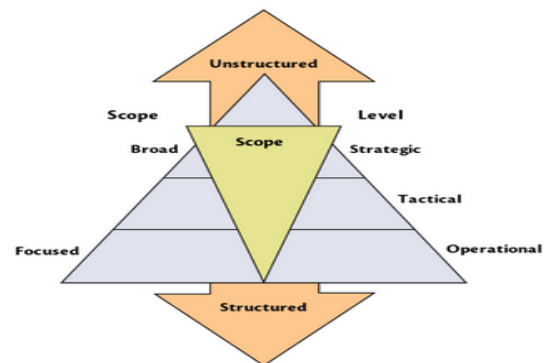


Figure 1 – Hierarchical Decision Levels [2]

The figure reveals that the strategic decision level covers a broad scope of unstructured problems while operational level decisions are more focused and structured. This is a clear approach to show how healthcare management decisions can be organized. When strategic decisions are performed, decision makers focus on, for example, patient groups, rather than an individual patient which is the focus of operational scheduling decisions. Despite its development more than 50 years ago, the framework presented in [2] is still widely accepted in decision support systems (DSS) research, as demonstrated in [3]. By breaking down DSS research literature into the classical hierarchies, their work reveals that the majority of business problems in DSS design science research have focused on the operational level. In contrast, our E-HOSPITAL workbench combines all levels in one digital platform.

Stakeholders in the Decision-Making

Figure 2 provides an overview of different stakeholders and their objectives, aiming to understand inefficiencies in hospitals, improve resource utilization, or to maximize profit.

We embed multiple mathematical models and its solution approaches from the literature to support these objectives in an integrated decision-making environment. End users such as hospital administrators, healthcare analytics specialists and other decision makers can use the proposed workbench to demonstrate/explore how mathematical models can improve resource planning and allocation decisions in hospitals. Furthermore, we illustrate the use of the workbench in a Continuous Improvement Unit (CIU) of a health board, described as a case study later in this paper.

Stakeholder		
Decision making level	Board of Directors of Care Providers	Nursing/ Operating Room Manager
Strategic	Resource planning decisions	Staffing decisions
Tactical	Identification of bottlenecks in shift or other tactical planning decisions	Shift scheduling
Operational	Identifying which elective patients may be responsible for increasing or decreasing profit margins for day-by-day planning	Appointment scheduling

Figure 2 – Stakeholders on Each Decision-Making Level

Related Work on Decision Support Tools

An early review of evaluation studies of clinical decision support tools in medical informatics was performed by Kaplan [21], while a more recent review focused on multi-morbid patients is provided by Fraccaro et al. [12]. More recently, Meulendijk et al. [22] presented a clinical decision support tool for physicians to optimize the patient's treatment plan and to avoid over-prescriptions.

Solving healthcare analytics and operations management problems in hospitals by means of a mathematical programming-based decision support tool has also been addressed in the literature. However, much more limited research is available as compared to decision support tools which focus on the clinical or medical perspective. In what follows, we provide an overview of, in our opinion, the four most relevant decision support tools that integrate DA, IT and OM for solving important and complex decision problems in healthcare delivery.

Joustra et al. [24] introduced a strategic decision support tool for patient mix decisions by enabling the management to alter the number of patients in various patient groups. Using sensitivity analysis, the impact of changing input parameters on key performance indicators can be studied. The authors presented a case study of the tool's application but did not provide details on its software implementation.

A tactical decision support tool for cyclic master surgery scheduling (MSS) implemented in Visual C++.NET was developed by Beliën et al. [5]. The system visualizes the impact of the MSS on the demand for various resources throughout the rest of the hospital. This system displays the impact of switching two physicians on the expected resource consumption pattern and additionally, it supports decisions made on the tactical level.

Another software system that was successfully applied on an operational decision level in a hospital is called ORSOS [9]. ORSOS is an enterprise-wide surgery scheduling and resource management system that automatically manages all surgical staff, equipment, and inventory using an engine that considers all of the clinical, financial, and operational criteria that must be addressed for each surgical event. Scheduling specific tasks, this tool supports decisions on the operational level.

Finally, Cayirli et al. [10] develop an appointment scheduling model that is also located on the operational decision level. It is implemented in an open-source online decision support tool and therefore not limited to a specific operating system.

We note that the systems which were published in the literature so far only support one of the three hierarchical decision-making levels, focusing either on the strategic, tactical or operational level. None of these applications integrate all three levels in one decision support tool. This ability will eventually allow opportunities to link solutions across the interfaces of these levels. To summarize, the main innovations of our E-HOSPITAL platform are two-fold: i) A unified, flexible and extensible workbench that combines different mathematical models of hospital resource planning problems by combining three classical hierarchical decision-making levels; ii) Formal, algebraic specifications of extensions of existing mathematical models are provided, implemented, and can be solved to optimality using sample instances, thus combining IT, operations and healthcare analytics in a single platform.

The remainder of this paper is structured as follows. We describe the workbench implementation and how we consider features that are highly relevant for practice by illustrating the use of the tool. Following this, we demonstrate the application of the workbench in a case study based on demand and capacity planning for hip fracture patients using real-world data from two hospitals. Lastly, we conclude with ideas for future work to extend the workbench, specifically highlighting opportunities linking the multiple levels.

Methods

When implementing the workbench, we focused on widely acknowledged theoretical concepts from the decision sciences literature that break down planning problems into different decision levels. When developing our modelling extensions, we incorporated practitioner's feedback into the existing models.

Implementation of the Different Decision Levels

Using the design objective of [2], seven approaches were selected from the literature that apply mathematical programming methods to provide decision support for healthcare OM problems. We also took into account the planning matrix of Hans et al. [19] who provided a similar classification of problems on the strategic, tactical and operational decision-making levels.

Strategic Decision Level

The strategic planning involves decision processes related to allocating resources, controlling organizational performance, establishing broad policies, and evaluating capital investment or merger proposals [26]. Decision support tools at this level need to help decision makers envision the future and negotiate with stakeholders by examining multiple scenarios [26].

These analyses are exactly what our workbench is aiming to provide: On the strategic level, Busse et al. [8] and Blake et al. [6] were selected. Both papers decide on the case mix of patients in hospitals while capacity constraints are considered. The difference between the two models is that Blake et al. [6] had target levels of physicians for treating patients and target revenue of the hospital, among others. In contrast, Busse et al. [8] followed an aggregate planning level to decide how many cases a hospital can support, given constrained resources. As a consequence, analyses can be run such as: Given operating room and bed capacity, what is the feasible number of patients to be treated within hospital budget limits?

Another scenario analysis is to examine the impact on revenue and the number of patients to be treated, given an increase or decrease in capacity.

Tactical Decision Level

Our workbench's tactical decision level consists of the tactical admission problem devised by Hans et al. [27]. Moreover, we include Master Surgical Scheduling (MSS) problems into that decision level, selecting the approaches of Blake et al. [7] and [25]. The difference between the two MSS papers is that Van Oostrom et al. [25] incorporate uncertainty into the planning while the approach in Blake et al. [7] is entirely deterministic.

Operational Decision Level

On the operational decision level, the operational shift scheduling problem in Beaulieu et al. [4], as well as an extension of the hospital-wide patient flow problem in Gartner et al. [14], were implemented.

Model Extensions

Before implementing the different models, we extended them to improve their applicability. On the strategic level, we extended the work of Busse et al. [8] on a temporal dimension. This allows users to insert expected values for different time periods for demand broken down by different diagnosis-related groups (DRGs). Another extension was the tactical planning problem of Vissers et al. [27] in order to capture demand for physical therapists and therapy rooms in the admission planning of patients. On the operational planning level, we extended the model of Gartner et al. [14] in order to capture admission decisions of patients, among others. The extensions are described in more detail on the workbench's repository: <https://github.com/drdanielgartner/ehospital>.

An Illustration of the Workbench

Figure 3 provides a specific example of the digital workbench. As can be seen, it separates the strategic, tactical and operational decision level using three tabs that are arranged vertically in the graphical user interface (GUI). Then, in each of the different planning levels, tabs are arranged horizontally, which separate the different approaches from each other.

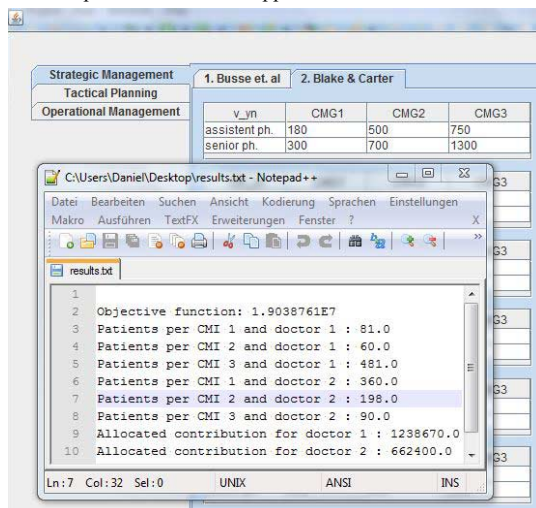


Figure 3 – Strategic Planning [6]

The illustration selects and solves the case mix planning problem of Blake et al. [6]. The GUI shows pre-specified default values e.g. for the number of case mix groups desired for each physician or the hospital capacity (e.g. beds and

operating room time). After solving the problem instance, the user can store the output in a text file, which provides information about the generated solution which also includes the objective function value and the cases assigned to each physician.

Installation Requirements

Before running the .jar file of the platform-independent environment which, again, can be downloaded at <https://github.com/drdanielgartner/ehospital>, IBM ILOG CPLEX [20] has to be installed. Also, at least version 6 of the Java Runtime Environment has to be installed.

Results

In this section, we describe how we incorporated a capacity planning model into the platform and how we carried out an analysis for a real-world project with a health board in the U.K.

Incorporating Capacity Planning into E-HOSPITAL – A Case Study

The objective of the case study is to show how the E-HOSPITAL workbench can be extended and used to support a real-world decision-making scenario. The task is to determine the optimal level of operating room and bed resource capacity required for treating hip fracture patients in a multi-hospital site in the United Kingdom. This problem is located at the strategic planning level. Rather than deciding on a narrow scope i.e. on individual patients at the operational level (e.g. patient scheduling decisions [18]), we decided on a broader scope, which is less structured and constrained [6]. Additionally, Busse et al. [8] models seem at first glance to be highly suitable. However, the board of directors who will use the decision support tool in future needs to determine the resource capacity level rather than the optimal number of patients given fixed capacities. Also, the board had specific usability requests e.g. to vary patient demand and length of stay.

Research Questions

The research questions which can be broken down into analytics and services planning are as follows:

Analytics-focused research questions

- How many patients require the service during a one-year planning horizon?
- What is the length of stay distribution of patients requiring hip fracture treatment in each of the hospital's catchment areas?

Strategic planning questions

- Fixing the catchment areas to the hospital sites, what are the total amounts of operating room time and bed capacity required?
- Pooling hospitals, what are the resource requirements for each of the hospitals?

Project Phases and Timeline

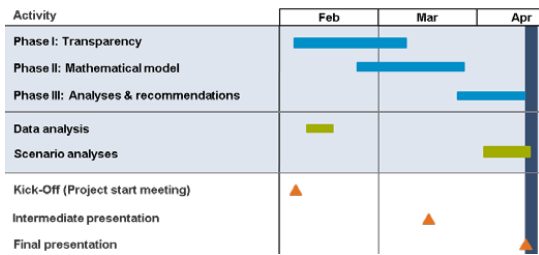


Figure 4 – Hip Fracture Demand and Capacity Planning

When carrying out the case study, we broke this project down into different phases as shown in Figure 4. In what follows, we will provide more details for each of the different project phases.

Transparency

In the first phase of the project which we called the “Transparency Phase”, we evaluated the length of stay (LOS) distribution because, in healthcare delivery, this is a major source of uncertainty and costs. Our data analysis revealed that the two hospitals that we studied (henceforth denoted as hospital 1 and hospital 2) are faced with a large inter-quartile range of LOS. Moreover, the median LOS is 28 days for hospital 1 and 23 days for hospital 2.

A more detailed analysis of the LOS data using histograms and Gaussian Kernel Density Estimators (KDEs) is shown in Figure 5. It reveals a left-skewed shape of the LOS distribution, which is similar to LOS distributions that can be observed in previous research [17, 23].

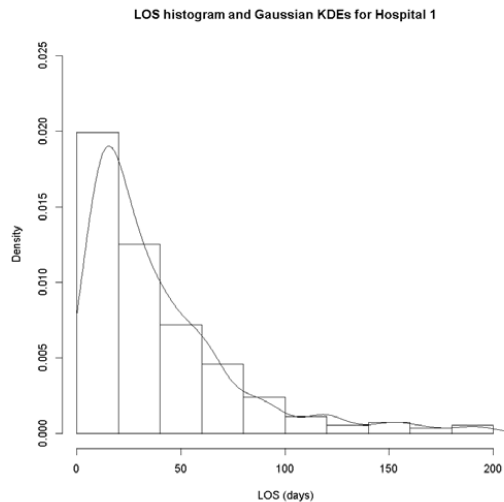


Figure 5 – Boxplots of LOS Distribution for Hospital 1

Mathematical Modelling

In the mathematical modelling phase, we used a model which is available in the workbench’s github repository. The model was developed in collaboration with Orthopaedic physicians and the GUI in collaboration with the physicians and the Modelling Lead of the Aneurin Bevan Continuous Improvement Unit (ABCI). The result is shown in Figure 6.

The upper part of the workbench revealed that patient demand reached 271 and 278 patients in the catchment area of hospitals 1 and 2, respectively, with the median LOS at 28 and

23 days. Manipulating the slider below the “#Patients” label and the slider below the “LOS quantile”, we observe that, for example, we can run our analysis for up to 50% more patients as compared to the baseline demand. Also, we can select any quantile for the LOS distribution. This reflects risk sensitivity for practitioners while ensuring that enough bed and operating room capacity is determined by the mathematical model since demand is fluctuating.

Assumptions, Analyses & Recommendations

For our analyses, we assumed that the average duration of a hip fracture surgery is 2.5 hours. To determine the demand, we selected patients admitted to the Accident and Emergency Unit (A&E) in 2014 and patients who were discharged from the hospital in 2014. We set up two scenarios as follows: Scenario 1 consisted of a run where we used the median (50% quantile) for length of stay. Also, we focused on actual patient demand observed in 2014. Moreover, we ran the model with a fixed assignment of patients to hospitals. This means that patients who arrive from hospital 1’s catchment area are exclusively treated in that hospital. The same holds true for hospital 2. In the second scenario, we include a third hospital (hospital 3) which will be built in the near future within the health board. In this scenario, the objective is to level bed capacity.

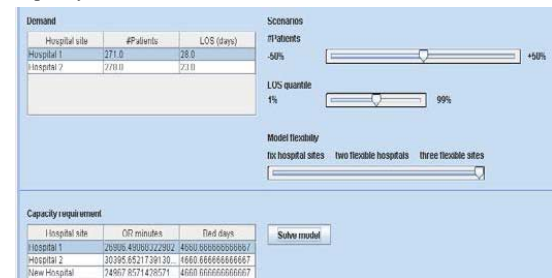


Figure 6 – Model integration in the E-HOSPITAL workbench and results of the fully-flexible model

The results of the scenario analysis revealed that using the fixed model, approximately 7,588 and 6,394 bed days are required for hospital 1 and 2, respectively. The results using the flexible model for three hospital sites (Figure 6) revealed that 4,661 bed days are required for each of the hospital sites. However, one can observe that the operating room capacity is different across the hospital sites which is attributed to the different patients’ LOS. Fewer patients are admitted to hospital 1, but have the same total bed days due to their longer LOS, but lower total OR capacity requirement.

Discussion

Compared with the current state of the art, the proposed platform can be considered as the first that unifies multiple models in one platform and extends it to increase the acceptability in health care. Another contribution that extends current state of the art is that multiple decision levels can be tackled by using this platform. One limitation, however, is that the commercial solver CPLEX has to be installed with the platform.

In the scenario analysis that we provided in the results section, we employed realized patient demand as a predictor for future demand. In other inpatient settings and especially for elective patients, the size of waiting lists has to be accounted for as well. Also, there are many more factors that determine length of stay such as quality of care, hospital discharge policies, and so on. However, many of these can be incorporated as site-

specific parameters into the mathematical models and solved for varying scenarios of parameter values.

Conclusion

In this paper, we described the development of a unified digital workbench for hospital resource planning that is based on a well-accepted, multi-level decision-making framework. The platform leverages information technology, operations management, and data analytics to support not only healthcare decision makers but also healthcare analytics and information systems specialists as well as educators of these topics. The tool combines the three classical hierarchical decision-making levels in one integrated environment. At each level, several decision problems can be chosen. Extensions of mathematical models from the literature are presented and incorporated into the workbench. In a case study using real-world data, we demonstrated how we used the workbench to inform capacity decisions in a multi-hospital site.

Future work will address the intersection between the different decision layers. Although the intersection between the strategic and the tactical layer have not yet been covered extensively due to computational complexity, our aim is to provide computationally tractable, heuristic methods to evaluate the intersection between multiple decision layers when optimal approaches are not feasible.

References

- [1] American Hospital Association. 2016. URL: <http://www.aha.org/advocacy-issues/medicare/ippa/index.shtml> (last retrieved: 15/12/2016).
- [2] Anthony, R.N. *Planning and Control Systems: A Framework for Analysis*, Division of Research, Graduate School of Business Administration, Harvard University, Boston. 1965.
- [3] Arnott, D., and Pervan, G. Design science in decision support systems research: An assessment using the hevrer, march, park, and ram guidelines. *Journal of the Association for Information Systems* **13** (2012), 923-949.
- [4] Beaulieu, H., Ferland, J.A., Gendron, B. and Michelon, P.A. mathematical programming approach for scheduling physicians in the emergency room. *Health Care Management Science* **3** (2000), 193-200.
- [5] Beliën, J. and Demeulemeester, E. and Cardoen, B. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling* **12** (2009), 147-161.
- [6] Blake, J.T. and Carter, M. A goal programming approach to strategic resource allocation in acute care hospitals. *European Journal of Operational Research* **140** (2002), 541-561.
- [7] Blake, J. and Donald, J. Mount Sinai hospital uses integer programming to allocate operating room time. *INFORMS Interfaces* **32** (2002), 63-73.
- [8] Busse, R., Schreyögg, J., and Stargardt, T. *Management im Gesundheitswesen*, Springer, Heidelberg, 3rd edition. 2013.
- [9] Carter, J. Timing is everything in the OR. *Health Management Technology* **21** (2000), 80-81.
- [10] Cayirli, T., Yang, K.K. and Quek, S.A. A universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management* **21** (2012), 682-697.
- [11] Clegg, R., Hardy, C., Nord, W.R. *Managing Organizations: Current Issues*, Sage, London. 1999.
- [12] Fraccaro, P., Casteleiro, M.A., Ainsworth, J. and Buchan, I. Adoption of clinical decision support in multimorbidity: A systematic review. *JMIR Medical Informatics* **3** (2015), e4.
- [13] Gartner, D. *Optimizing Hospital-Wide Patient Scheduling*. Lecture Notes in Economics and Mathematical Systems, Springer, Heidelberg. 2015
- [14] Gartner, D. and Kolisch, R. Scheduling the hospital-wide flow of elective patients. *European Journal of Operational Research* **223** (2014), 689-699.
- [15] Gartner, D., Kolisch, R., Neill, D.B. and Padman, R. Machine learning approaches for early DRG classification and resource Allocation. *INFORMS Journal on Computing* **27** 2015, 718-734.
- [16] Gartner, D. and Padman, R. 2015. "Improving Hospital-Wide Early Resource Allocation through Machine Learning." MEDINFO 2015: EHealth-enabled health: Proceedings of the 15th world congress on health and biomedical informatics **216** (2016), p. 315.
- [17] Gartner, D. and Padman, R. 2016. Mathematical Modelling and Cluster Analysis in Healthcare Analytics - The Case of Length of Stay Management, International Conference on Information Systems, Dublin.
- [18] Gartner, D. and Padman, R. 2017. Handbook of Research on Healthcare Administration and Management, chapter Mathematical Programming and Heuristics for Patient Scheduling in Hospitals: A Survey, 168, IGI Global.
- [19] Hans, E.W., van Houdenhoven, M. and Hulshof, P.J.H. 2011. A framework for health care planning and control. Chapter 12 of R. Hall (Ed.), Handbook of Health Care Systems Scheduling, Springer, 168, 303-320.
- [20] IBM ILOG CPLEX. 2015. <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/>, (last retrieved: 11/30/2015).
- [21] Kaplan, B. Evaluating informatics applications - clinical decision support systems literature review. *International Journal of Medical Informatics* **64** (2001), 15-37
- [22] Meulendijk, M.C., Spruit, M.R. Jansen, P.A.F., Numans, M.E. and Brinkkemper, S. STRIPA: A rule-based decision support system for medication reviews in primary care. *ECIS 2015 Research-in-Progress Papers*. Paper 29.
- [23] Min, D. and Yih, Y. Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research* **206** (2010), 642-652.
- [24] Joustra, P.E., de Wit, J., Van Dijk, N.M. and Bakker, P.J.M. 2011. "How to juggle priorities? An interactive tool to provide quantitative support for strategic patient-mix decisions: an ophthalmology case." *Health Care Management Science* **14** (2011), 348-360.
- [25] Van Oostrum, J.M., Van Houdenhoven, M., Hurink, J.L., Hans, E.W. Wullink, G. and Kazemier, G. A master surgical scheduling approach for cyclic scheduling in operating rooms departments. *OR Spectrum* **30** (2008), 355-374.
- [26] Power, D. How does organization level impact decision support?, *DSS News* **9** (2008).
- [27] Vissers, J.M.H., Adan, I.J.B., and Bekkers, J.A. Patient mix optimization in tactical cardiothoracic surgery planning: A case study. *IMA Journal of Management Mathematics* **16** (2005), 281-304.
- [28] Zhang, Y. and Padman, R. 2015. "Innovations in chronic care delivery using data-driven clinical pathways. *American Journal of Managed Care*, special issue on Health IT, **21** (2015), e661-e668.
- [29] Zhang, Y. and Padman, R. 2016. Data-driven clinical and cost pathways for chronic care delivery. *American Journal of Managed Care*, special issue on Health IT **22** (2016), 294-298.

Addresses for correspondence

gartnerd@cardiff.ac.uk, rpadman@cmu.edu
and <https://github.com/dranielgartner/ehospital>

Consumer Health Informatics Aspects of Direct-to-Consumer Personal Genomic Testing

Kathleen Gray^a, Remya Stephen^a, Bronwyn Terrill^c, Brenda Wilson^g, Anna Middleton^f
Rigan Tytherleigh^{ab}, Erin Turbitt^b, Clara Gaff^{ae}, Jacqueline Savard^d,
Chriselle Hickerton^b, Ainsley Newson^d, Sylvia Metcalfe^{ab}

^aThe University of Melbourne VIC Australia, ^bMurdoch Childrens Research Institute, Melbourne VIC Australia, ^cGarvan Institute of Medical Research, Sydney NSW Australia, ^dUniversity of Sydney, NSW Australia, ^eWalter and Eliza Hall Institute, Melbourne VIC Australia, ^fWellcome Trust Sanger Institute, Cambridge UK, ^gUniversity of Ottawa, ON Canada

Abstract

This paper uses consumer health informatics as a framework to explore whether and how direct-to-consumer personal genomic testing can be regarded as a form of information which assists consumers to manage their health. It presents findings from qualitative content analysis of web sites that offer testing services, and of transcripts from focus groups conducted as part a study of the Australian public's expectations of personal genomics. Content analysis showed that service offerings have some features of consumer health information but lack consistency. Focus group participants were mostly unfamiliar with the specifics of test reports and related information services. Some of their ideas about aids to knowledge were in line with the benefits described on provider web sites, but some expectations were inflated. People were ambivalent about whether these services would address consumers' health needs, interests and contexts and whether they would support consumers' health self-management decisions and outcomes. There is scope for consumer health informatics approaches to refine the usage and the utility of direct-to-consumer personal genomic testing. Further research may focus on how uptake is affected by consumers' health literacy or by services' engagement with consumers about what they really want.

Keywords:

Access to Information; Consumer Health Information; Genomics

Introduction

Advances in healthcare have created a need for informed consumers. To meet this need, the Internet has provided a way for consumers to access information; including access to medical literature, connection with patient social networks and the creation of open health data. Notoriously, the Internet is also a massive source of misinformation about health. Consumer health informatics (CHI) is concerned with the health information structures and processes that enable people who are not clinically trained (so-called "consumers") to be informed for the purpose of managing their own health. Some consumer health information resources are purpose-built, such as health information literacy aids and personal health records, while others show technological appropriation, for instance using Facebook, Youtube or Twitter for health self-management.

From a synthesis of findings reported in recent reviews [1, 2], the focus of CHI is on the types of tools or methods that can make valid health data, information and/or knowledge resources available to consumers, and the aim of CHI is to understand and improve the ways that these tools or methods:

enable access, materially and intellectually, by consumers; address the health needs, interests and contexts of consumers; allow direct interaction by consumers without the presence of a healthcare professional; personalise and / or socialise consumers' interactions about their health needs and interests; aid consumers' health self-management and/or self-reported outcomes; facilitate consumers' engagement in clinical diagnosis and/or treatment. This paper uses these CHI considerations to explore one relatively new health-related phenomenon on the Internet.

Health-related personal genomic testing services have been available directly to consumers over the Internet for about ten years. Consumers register with a service, create a user account and make a payment online, then use a postal or courier service to ship some saliva or other body tissue to a laboratory, and generally receive their test results and interpretive information by email. These services are increasing in number and reach [3]. Currently over 130 service providers advertise to consumers in the English language; many of them do not require a clinical referral and are priced cheaply by comparison with clinically mediated testing [4]. Popular uptake is rising. One service, Mapmygenome (www.mapmygenome.in), according to its publicity materials, aims to touch 100 million lives and save a million lives by 2030.

However, there are mixed views as to the benefits and risks of direct-to-consumer personal genomic testing (DTC-PGT) – for example, whether empowerment or overdiagnosis is more likely to be its net result [5]. In Australia the direct sale by company of a direct-to-consumer personal genome test for health information is prohibited. However, many companies in Australia have adopted a model where they offer a personal genome test such that the sample is processed overseas. Further, Australian regulations do not stop consumers from arranging online for tests to be done outside their country of residence. Health authorities can only caution consumers about their use [6].

The function of DTC-PGT as a form of consumer health information is acknowledged by the US National Library of Medicine [7] and its use as such has been investigated in controlled settings [8]. An opportunity to use a CHI lens to form a clearer view of the information structures and processes that typify DTC-PGT using real-world data has arisen through research in a multi-disciplinary, multi-stage study to explore broadly Australians' expectations of personalised genomics. The Genioz (Genomics: National Insights of Australians) project (www.genioz.net.au) which began in 2015 involves focus groups, a quantitative survey, semi-structured interviews and ethical critique.

This paper specifically reports research into DTC-PGT services available to Australian consumers, and into public ideas about these services that have emerged in GenioZ focus groups. Our aim here is to use CHI as a framework to explore whether and how DTC-PGT can be regarded as a form of information which assists consumers to manage their health.

Methods

We used two data sources. First within the broad definition of DTC-PGT provided to consumers by the US National Library of Medicine [7], we selected web sites to represent a cross-section of services offering wellbeing, disease, individual and ancestral information, including some market leaders and some less well known companies. We analysed web site content that was publicly accessible in late 2015 from 10 different providers, on a total of 69 web pages: AncestrybyDNA, DNA Worldwide, EasyDNA, FamilyTreeDNA, GenetrackAustralia, GenomicsforLife, GTLDNA, Mapmygenome, PathwayGenomics, 23andMe. Data from web sites is not identified by company name, so as to avoid potential perception of bias.

Then we analysed the transcripts of seven age-category stratified GenioZ focus groups in two capital cities in 2015, involving a total of 56 members of the public. People were recruited regardless of whether they had ever undergone personal genomic testing; purposive sampling was done for gender and age. Apart from individuals' gender and age details, data are de-identified to protect confidentiality.

Since these datasets were not generated specifically to explore consumer health informatics aspects of DTC-PGT, we took a summative approach to analysis, that is, identifying certain content with the purpose of understanding it in a particular context [9]. We used a coding guide (Table 1) to identify content, a framework method [10] to chart its occurrence, and an abstract level of interpretation [11].

The coding guide supported a junior researcher to perform a systematic search for terms associated with managing data, information, and/or knowledge (manifest content); these three fundamental information science concepts align with formal models of health informatics [12, 13] but are often used interchangeably in everyday language. Then the latent content of the charted occurrences was analysed by a senior researcher with consumer health informatics expertise, in the context of the CHI concepts and concerns summarised in the Introduction section of this paper.

Table 1 – Example Terms used in Manifest Content Analysis

Data Management	Information Management	Knowledge Management
Administer	Categorize	Apply
Analyse	Classify	Decide
Capture	Filter	Learn
Integrate	Organise	Reflect
Monitor	Search	Synthesise
Store	Select	Understand

This study is preliminary to a more in-depth analysis of a larger dataset from the GenioZ project. In a larger-scale and more nuanced study, quantification of manifest content will have greater significance and other Smethods, such as sentiment analysis, also may be illuminating.

For the purpose of the present exploratory study the Results section of this paper provides selected examples only of manifest content analysis, and the Discussion singles out noteworthy results of latent content analysis.

Results

Resources and Services Offered by DTC-PGT Providers

Data is the primary resource offered to consumers of PGT services, in the form of reports of results from tissue sample testing. These reports may cover a wide spectrum of aspects of health, wellbeing, fitness and identity, based on a variety of markers, accompanied or not by raw data files. A summary of what is offered on the ten provider web sites (represented by numbers 1 to 10) is as follows:

1. Genetic information categorized four ways: carrier status, wellness, traits and ancestry. Over 100 health conditions and traits. Provides raw data.
2. Paternity, relationship, prenatal, ancestry, clinical, other tests (genetic fingerprinting, DNA profiling), health tests (skin, children's DNA, diet, nutrition, fitness, wellness, weight and lifestyle), animal tests. Based on 21 markers. Provides raw data.
3. 50 physiological and lifestyle traits, health conditions and inherited conditions. Personal, brain wellness, TB diagnostics, molecular diagnostics, forensics, merchandise. 16 markers. Provides raw data.
4. Paternity, immigration, ancestry, extended family, drug and alcohol, pets. 46 markers.
5. General health and wellness, liquid biopsy, hereditary cancer, pharmacogenomics, carrier screening. A variety of conditions including cancer risk, cardiac health, inherited diseases, nutrition and exercise response, as well as drug response for specific medications including those used in pain management and mental health. 75 markers. Provides raw data.
6. DNA origin, DNA world view, paternal and maternal. Raw data available is limited. 144 markers.
7. Disease screening, prenatal screening, paternity, immigration, > 700 inherited diseases.
8. Family finder, father's line (37, 67 or 111 markers), mother's line. Raw data provided in CSV or XML format.
9. Paternity, maternity, sibling, grandparentage, aunt/uncle, twin, ancestry test. 91 markers.
10. Paternity, prenatal, relationship, ancestry, other tests (infidelity DNA testing, DNA profile, surrogacy DNA testing, additional services, semen detection DNA testing service, genetic reconstruction testing, non-invasive prenatal test for Down syndrome). 21 markers.

Some data management features are explicit on some web sites, such as these examples of access, security, and visualization:

- "choose to authorise [the provider] to share their data with specific individuals" (1)
 "access an interactive version of your report anytime, anywhere" (5)
 "once the testing is completed and the results provided to the client, all raw data is purged from our system" (6)
 "browser allows you to see where on your chromosomes a match shares DNA with you" (8)

Information services that contextualise results reporting may be available. Selected examples are:

- "reports in four categories: carrier, ancestry, wellness and traits" (1)
 "interesting facts about your own haplogroup.... access to information about your extended genetic cousins" (2)
 "an expert review of your case looking at possible reasons for the results" (4)
 "a detailed manual that explains your results" (6)

“an in-depth guide to understanding and interpreting your results” (8)

Information filtering also appears, in the form of messages advising on appropriate use of the data. For example, seven providers’ web sites included content that recommended against using their genetic tests for clinical or diagnostic purposes.

Some providers socialise the sharing of information by hosting user comments. For example,

“very interesting, but I’m a little unhappy that the information provided for conditions genetic genealogy and inter-species genealogy is so poor despite a fantastic database, genbank and published tables for much of this stuff” (1, user comment)

Knowledge management facilitation for users is promoted prominently. Providers’ web site content uses the language of insight into self and others, actionable information, decision-support. For example:

“knowing your level of genetic predisposition toward developing particular conditions can help both you and your children live healthier and longer lives” (2)

“help you understand what your genetic variations mean and how it may affect you. ... learn about inherited and acquired genetic health risks” (3)

“reports provide actionable information and recommendations that provide individuals with a roadmap to achieve optimal heart health” (5)

“discover your geogenetic links by seeing your match with studied populations and anthropological groups around the world” (10)

Several providers offer access to online genetic counselling or other clinical services, as a way for consumers to leverage test results and information. Knowledge management may also include suggestions on engaging expertise, for example:

“empowers the physician to help patients make informed decisions regarding diet and exercise, while also providing information about medication response” (5).

Communicating with other service users or other communities of interest is sometimes an option, for example:

“enhance your experience with relatives” (1)

“names and emails of your matches ... are provided to you in order to collaborate” (8).

Ideas about DTC-PGT Held by Members of the Public

Examples of considerations that members of the public (Participant ID number / Gender / Age) voiced about managing data, information and knowledge are provided here. Key data management concepts were access and governance.

- The right to data (restricting citizens’ access to information was the way one person interpreted one DTC-PGT provider’s legal issue over non-compliance with laboratory testing regulations):
“the authority in the [United] States said ‘no you can’t do that’, they didn’t feel it appropriate that people should know this information because they may misunderstand it which is rather strange, there is a big battle going on at the moment and it looks like they are going to get the okay to release that information” (P40/M/72)
- Direct and timely access to the data:
“go online and find this stuff and just put an order in, pay for it and get it delivered to our door” (P5-6/F/36)
“you get massive bottlenecks ... now you’ve got the thousand dollar human genome and the twenty dollar bacterial genome ... why they don’t just sequence things in the hospital more often” (P23/M/22)
- Management and governance of the data:

“I’m only interested in looking at myself, I don’t like it to be publicly accessible or set up as a database” (P12/M/32)

“they [test providers] might keep a database of all the genetic information about you, for future reference” (P13/F/25)

“the legitimacy of the company that is handling it, what might they do with it” (P14/F/20)

“who are you going to offer these tests, how are you going to dictate what information is given out on the test, how are you going to ensure the validity of these tests” (P23/M/22)

People expressed attitudes about information management issues, particularly usefulness and shared use:

- Information usefulness limited in scope:
“this genetic testing gives you a window into what the possibilities are, but it doesn’t define you” (P20/M/23)
“if I paid my hundred dollars today I don’t know that it’s actually better than what it might be if I survive another ten years, if I’m getting imperfect information how useful is the information” (P23/M/22)
 - Information usefulness contingent on external validation:
“find other people who have done the test ... a while ago and then see if the things that came back from their test ... what good are they, the test[s], a certain number of years down the track” (P7-6/M/22)
“if the website looked really legit ... if the government was recommending it ... if other people say this genetic testing has really helped me ... it depends on the person [delivering the message] and how they interpret it for you” (P25/M/21)
 - Information usefulness linked to clinical consultation:
“if having an understanding of genetic make-up is providing information that shows patterns in or information about it will help health conditions and that can then inform medical practitioners to come up with preventative measures” (P5-1/M/41)
“if my doctor said I would gain a lot from this I would do it” (P25/M/21)
 - Information uses shared with other parties, not necessarily aligned with the consumer’s values:
“people who want to try and genetically dictate what their children’s genes are going to be” (P14/F/20)
“that information would be really useful for me ... but for my health insurer, that’s an extra risk for them” (P53/M/42)
- Knowledge management attitudes of members of the public were mixed.
- Self-knowledge for its own sake, sometimes with unrealistic expectations:
“sometimes just knowing can make things a lot easier and having more time to understand it and understand what’s going to happen and ... knowing a timeline, it’s almost like knowing this is what I have before this is going to start setting in” (P7-1/F/24)
“for a company to say, we can decode it and tell you with any certainty ... psychologically, it could be wish fulfilling” (P6-2/F/63)
“if I did find out that I’m going to have an issue with my health in the future, I wouldn’t want them to tell me when it was going to be, I wouldn’t want to put a timer on myself” (P7-6/M/22)

“somehow that would give you some idea of how long you’re going to live for” (P49/M/77)

- Knowledge for the sake of significant others:
 “for the benefit of my next generation, yeah, and I can afford it, I will definitely go for that. To find out if there’s anything which can be mended before [conception / birth]” (P5-4/F/45)
 “whilst I don’t want my information, I really don’t want to know where my life’s going to go to, I sort of feel that for my child or a grandchild or another person, I would do that and find out because it’s something that can be prevented” (P46/F/63)
 “we’ve got a social responsibility ... to find out whether we’ve got [conditions that could be passed on to children]” (P47/F/75)
- Knowledge as a prompt to action; or a deterrent:
 “if I knew I had a predisposition to melanoma, I wouldn’t be going out in the sun. It would change how I do stuff in my life, some genetic testing could change your behaviour” (P24/F/23)
 “I considered whether or not I got tested to see if my body type was one that would gain muscle quicker than someone else ... I considered it but then ... if I did find out that ultimately I’m a really fat body ... I’d give up on life ... like ‘well there’s no point’ ... and then ... the course of your life changes” (P25/M/21)
 “if you knew there was a tendency to diabetes in your inherited characteristics you might pay a bit more attention to your diet as you age” (P28/F/72)
- Managing uncertainty or bad news:
 “if you think you could get a report that you could get breast cancer, then you spend the next x number of years with sleepless nights thinking ‘oh am I going to get breast cancer or not, is today the day?’ ... that sort of uncertainty is a little bit of a drawback” (P20/M/23)
 “anybody can get their genome sequenced for a thousand dollars, and it means nothing ... we have no idea what the mutations mean, we have nothing to compare them to overall, we have no baseline and no comparator, so that’s cool, you may have really good genetics, you may not” (P23/M/22)
 “what that might allow me to learn about myself that I might or might not like ... you’d measure your own genome with a view to find out something as I say that you might or might not like” (P58/M/57)

Discussion

Within the overall focus group dataset, comments that could be related to management of DTC-PGT data, information and knowledge occurred infrequently. This is not so surprising since the general aim of the data collection was to explore broader public attitudes and understandings regarding personal genomics. The variety of ideas that was put forward suggests that there is scope for more specific investigation on this topic. People appeared to be unfamiliar with the specifics of DTC-PGT data reports and additional information services. Major concerns about privacy did not emerge from these groups. People had more to say about the aids to knowledge that they thought DTC-PGT could provide, and some of what they said was similar to the benefits described on provider web sites. Some expectations were inflated, for example, the idea that you can learn how long your life will be (though indeed some companies suggest that this is possible).

Turning to the broader question of how DTC-PGT can function as consumer health information, the findings here provide evidence that it has some of the features that would be advocated by CHI, and that it offers some scope for the kinds of improvements that could be guided by CHI.

In terms of features, its essential rationale is to allow consumers direct interaction with data, without the presence of a healthcare professional. It appears to enable access by consumers, both materially (people did not foreground connectivity or cost or other issues of accessibility) and intellectually, considering the information services that providers offer to explain testing and support reading of test results. Personal genomic test results by their very nature personalise consumers’ interactions with information about health, to the extent that they identify the individual’s inherited conditions, risk factors and potential response to therapies. Likewise, the fundamental connection between the individual’s test results and their genetic family members’ past, present and future socialises this form of consumer health information; further, some providers offer suggestions and facilities for a consumer to exchange data and information with family members and other consumers. A few DTC-PGT providers also encourage and support consumers to share data with a health professional, such as a genetic counsellor available through the provider, or their own treating clinician.

In terms of improvements, the validity of the data and the value of the surrounding service which DTC-PGT provides to consumers is a key unresolved aspect of whether it is fit for purpose as consumer health information. This is unresolved partly because the consumer’s personal genomic data is being reported and interpreted to them as a formulaic business transaction, within a dynamic field of knowledge where there is much still to be determined about the underlying biomedical science and social science. CHI could be applied to improve tools that assist with health literacy in this area and thus create more discernment in the consumer market for DTC-PGT and the after-market for medical services (as described in [14]). This is unresolved also because web site analysis shows considerable variation in the description of services, so that it is complicated for consumers to make choices about whether a service provider will meet their health needs and difficult for them to be confident about using the ensuing information as a basis for health decisions. Within a model of self-regulation of direct-to-consumer services, there is scope for CHI to contribute to formulating and monitoring industry standards (on data management, for example [15]) and to developing aids that allow consumers to compare the services on offer.

On the question of whether DTC-PGT is actually working as a form of consumer health information, findings from the focus groups suggest that it has not yet captured widespread public attention, certainly in Australia. By comparison, other more conventional forms of consumer health information provided by public agencies (such as Better Health Channel) and private interests (e.g. BUPA Health and Wellness) are much more familiar to the Australian public. It is noteworthy that DTC-PGT has not taken off given high rates of Internet use (around 80% of Australians were using the Internet to look for health information nearly a decade ago [16] and 61% had purchased goods and services online during a three-month period surveyed in 2014-15 [17]).

This suggests that DTC-PGT as a model of consumer health information on the Internet may not be working in terms of the way it addresses two particular CHI concerns. Does it address the health needs and interests and contexts of consumers? People had no direct experience of working with DTC-PGT information to report. They had ambivalent ideas about whether it could do this – some expressed personal curiosity or a sense

of obligation to family members, but others said that they preferred not to know or that they wanted more assurance about the information quality. Does it help consumers with their health self-management decisions and outcomes? Some people hypothesized that it would be an effective tool to support healthy behaviours and constructive health actions including seeking professional help. Others were less certain, alluding to anxiety and unhappiness that could ensue.

What consumer health information works and why is often explained by consumers' health literacy, or by service providers' engagement with consumers about what they really want [18]. Both of these factors in DTC-PGT, and possibly others that were not identified in this study, deserve further research.

Conclusion

This paper presents an exploratory investigation of whether we can regard DTC-PGT as a form of consumer health information, in actuality or potentially. In actuality, it does not seem as widespread or well-known among the Australian public as one might expect from industry reports. Its potential as consumer health information is unclear overall. We found inconsistent offerings and divergent attitudes, and this situation is echoed in other places. On one hand a highly articulate consumer feels "seduced by the idea that it [information from PGT] fills a void in unanswerable questions, yet without enough knowledge to properly interpret its true ramifications" [19]. On the other hand a highly regarded physician argues that health and consumer authorities should put greater effort into working with reputable consumer genomics companies to allow them to deliver more health information [20].

DTC-PGT is an Internet phenomenon which has been the subject of considerable research, but perhaps not enough through the lens of consumer health informatics. Based on the exploratory study reported here, the question of whether DTC-PGT can be or should be taken seriously as a form of consumer health information cannot be resolved. However, the need to address this issue spiked sharply from April 2017 when, after years of deliberation, the US Food and Drug Administration finally made it legal for one DTC-PGT company (23andMe) to market genetic health risk tests for a limited number of conditions [21]. Whether and how the forms and functions of health information of this kind can be systematically made safe, effective, accessible and equitable will have a major influence on the DTC-PGT industry and on healthcare generally.

From a CHI perspective, that is, seeking to understand how the Internet may be used optimally to increase the reach and sophistication of high quality information, DTC-PGT surely warrants further research and development before its growing range of information products and services can be either dismissed or endorsed as aids to health self-management.

Acknowledgements

The work reported in this paper was funded in part by the Australian Research Council (DP150100597). Additional work was undertaken as a student research project in the University of Melbourne Master of Information Technology (Health) degree.

References

- [1] G. Demiris, Consumer health informatics: Past, present, and future of a rapidly evolving domain, *Yearb Med Inform* **25** (2016), (Suppl. 1).
- [2] D. Flaherty, L. Hoffman-Goetz, and J.F. Arocha, What is consumer health informatics? A systematic review of published definitions, *Inform Health Soc Care* **40** (2015), 91-112.

- [3] Global Industry Analysts, Inc. *Direct-to-Consumer (DTC) Gene Testing: A Global Strategic Business Report*, 2012. http://www.prweb.com/releases/DTC_genetic_testing/direct_to_consumer_tests/prweb9780295.htm
- [4] A.M. Phillips, 'Only a click away' - DTC genetics for ancestry, health, love...and more: A view of the business and regulatory landscape, *Appl Transl Genom* **8** (2016), 16-22.
- [5] L. Covolo, S. Rubinelli, E. Certti, and U. Gelatti, Internet-based direct-to-consumer genetic testing: A systematic review, *J Med Internet Res* **17** (2015), e279.
- [6] Australia Department of Health, *National Health Genomics Policy Framework 2017-2020 Consultation Draft*. 2016. <https://consultations.health.gov.au/genomics/national-health-genomics-policy-framework>
- [7] US National Library of Medicine, *Genetics Home Reference: What is direct-to-consumer genetic testing?* 2016. <https://ghr.nlm.nih.gov/primers/testing/directtoconsumer>
- [8] O. Shaer, O. Nov, J. Okerlund, M. Balestra, E. Stowell, L. Ascher, L., et al., Informing the design of direct-to-consumer interactive personal genomics reports, *J Med Internet Res* **17** (2015), e146.
- [9] H.-F. Hsieh and S.E. Shannon, Three approaches to qualitative content analysis, *Qual Health Res* **15** (2005), 1277-1288.
- [10] N.K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, Using the framework method for the analysis of qualitative data in multi-disciplinary health research, *BMC Med Res Methodol* **13** (2013), 117.
- [11] M. Vaismoradi, H. Turunen, H., and T. Bondas, Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study, *Nurs Health Sci* **15** (2013), 398-405.
- [12] A. Georgiou, Data, information and knowledge: The health informatics model and its role in evidence-based medicine, *J Eval Clin Pract* **8** (2002), 127-130.
- [13] C. Ronquillo, L. Currie, and P. Rodney, The evolution of data-information-knowledge-wisdom in nursing informatics, *ANS Adv Nurs Sci* **39** (2016), E1-E18.
- [14] J.L. Krieger, F. Murray, J.S. Roberts, and R.C. Green, The impact of personal genomics on risk perceptions and medical decision-making, *Nat Biotechnol* **34** (2016), 912-918.
- [15] L.I. Laestadius, J.R. Rich, and P.L. Auer, All your data (effectively) belong to us: data practices among direct-to-consumer genetic testing firms, *Genet Med* (2016), doi:10.1038/gim.2016.136.
- [16] D. McDaid, and A.L. Park, *Online Health: Untangling the Web - Evidence from the BUPA Health Pulse 2010 International Healthcare Survey*, London, BUPA, 2011. http://www.bupa.com.au/staticfiles/Bupa/HealthAndWellness/MediaFiles/PDF/LSE_Report_Online_Health.pdf
- [17] Australian Bureau of Statistics, *8146.0 - Household Use of Information Technology, Australia, 2014-15*, 2016. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0>
- [18] L. Kayser, A. Kushniruk, R.H. Osborne, O. Norgaard, and P. Turner, Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs, *JMIR Hum Factors* **2** (2015), e9.
- [19] P. Fall, Waiting for cancer. *New York Times*, 2016, September 1 www.nytimes.com/2016/08/31/well/live/waiting-for-cancer.html
- [20] E. Topol, quoted in H. Mack. In depth: consumer genomics and the road to legitimacy. *MobiHealthNews*, 2016 November 23 <http://www.mobihealthnews.com/content/depth-consumer-genomics-and-road-legitimacy>
- [21] FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions, 2017 April 6 <https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm>

Address for correspondence

Kathleen Gray kgray@unimelb.edu.au

Designing a Mobile Health Application Prototype for the Management of Interstitial Cystitis/Painful Bladder Syndrome

Janessa Griffith^{a,b}

^a Women's College Hospital Institute for Health System Solutions and Virtual Care, Toronto, Ontario, Canada

^b Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada

Abstract

The design of an early mobile health application (app) prototype to manage interstitial cystitis/painful bladder syndrome, a chronic condition characterized by recurrent pain/discomfort in the bladder and pelvic floor, is described. The purpose of this app prototype is to help people who have IC/PBS manage and learn what triggers their symptoms. Another aim of this research was to provide an example of how sex and gender could be included into the design of a health information system. Based on a literature search of common symptoms and challenges faced by people living with IC/PBS, the researcher created an app prototype design including many features: resources for relaxation, mental health, intimacy, pregnancy, and daily life; reminders for appointments, and medication; logs for diet, activity, sleep, pain, menstruation; and a link to a public washroom locator. This prototype will later undergo usability and content evaluation.

Keywords:

Telemedicine; Interstitial Cystitis; Medical Informatics

Introduction

Interstitial cystitis/painful bladder syndrome (IC/PBS) is a chronic condition and those who have it experience recurrent pain or discomfort in the bladder and pelvic floor in the absence of an infection/disease. Pain/discomfort, urgent and frequent need to urinate and nocturia are common symptoms of IC/PBS [1]. Some people experience worsened symptoms with particular foods, activities, and menstruation [1]. The disease is more prevalent in the female population, with approximately 3 to 8 million (2.7% to 6.53%) female adults over 18 years of age [2] and 2 to almost 5 million male adults (1.9% to 4.2%) between 39 and 74 years of age [3] in the USA experiencing IC/PBS symptoms.

Despite the prevalence, no app is available to manage IC/PBS symptoms. The goal of this research is to design an app prototype for IC/PBS management and to help those with IC/PBS learn what triggers their symptoms.

Another goal of this research was to demonstrate how sex and gender considerations could be incorporated into health informatics solutions to support inclusivity. For example, as people with IC/PBS who are female experience worsened symptoms during menstruation, considerations will be built into the prototype. Another study focusing on people with IC/PBS who are female found that participants voided more than 30 times per day during a flare, impacting all areas of their lives [4]. A flare refers to a period of increased or worsening of

symptoms [5]. This may impact genders in different ways. For example, activities and responsibilities of people with IC/PBS may differ by gender (e.g. caregiving or social activities), and this app prototype seeks to consider these needs through its features.

Methods

The researcher surveyed the Canadian sites of the two most popular digital media stores for iOS and Android using the search terms "Interstitial Cystitis," "painful bladder syndrome" and "IC/PBS." No apps were available to manage IC/PBS. Only one app was available for people living with IC/PBS to monitor and make decisions about their diet and one app providing information. Based on a literature review on symptoms associated with IC/PBS [1-4, 6, 8-12], the researcher developed features for an app to help people with IC/PBS manage their chronic condition and learn to recognize what may trigger a flare. Specifically, the researcher identified common symptoms and patient concerns associated with IC/PBS and designed features aimed to help people manage this chronic condition. The researcher also identified ways in which the experiences of living with IC/PBS differed by sex and gender and tried to design features that would make the app more inclusive to these unique needs. Physicians and patients were not consulted for this preliminary design; however, they will be included for future iterations.

This app prototype represents the preliminary design, whereby features were drafted based on findings from the literature. Future iterations will include a user-centred design. A user-centred design was not explicitly used in this preliminary iteration because the researcher was more concerned with identifying what was already known in the literature and also due to budgetary constraints.

Based on these features, the researcher created a prototype design using Proto.io, online prototyping software.

Results

- The following features were included in the app prototype design:
- **Information:** The app prototype includes links to evidence-based websites where patients can learn about IC/PBS. Information is also tailored according to the sex, gender, and age of the user. For example, a user who identifies as being male would be directed

to physiotherapists who specialize in the male pelvic floor.

- **Relaxation:** The American Urological Association developed guidelines for managing IC/PBS and suggested several treatments: relaxation/stress management, pain management, patient education, and behaviour modifications [3]. Thus, the app prototype includes links to relaxation techniques.
- **Food and beverage log:** Another feature is a calendar diary where users enter the food and beverages they consumed, which corresponds to the pain log. This information may be helpful for users to present to their physicians to draw correlations between, for example, particular foods and symptom flares, and to support the development of a pain management plan. A study surveying participants on foods that worsened IC/PBS symptoms found that citrus, tomatoes, certain spicy foods, artificial sweeteners, coffee, tea, and alcohol were common triggers [6]. An app already exists to help people with IC/PBS select food that will not trigger their symptoms [7]. Thus, this app prototype includes a link for users to refer to this existing app.
- **Activity log:** Similar to the food and beverage diary, an activity log is included (Figure 1) which documents any exercise the user participated in, and events at particular times. This is also linked to the pain log. Thus, users can present this information to their care provider to identify potential activity triggers.

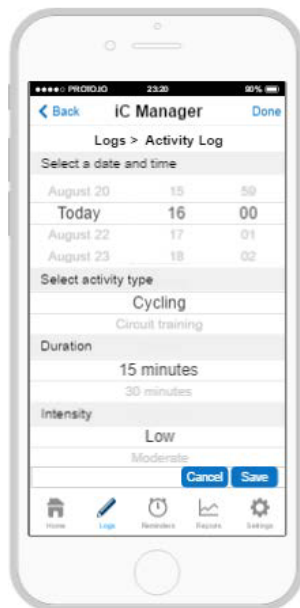


Figure 1– Activity log

- **Sleep log:** People living with IC/PBS may experience difficulty sleeping due to nocturia and other symptoms [8]. Therefore, a sleep log is included to capture any correlations between sleep disturbances, and, for example, diet, or physical activity.

- **Pain log:** People who have IC/PBS may experience pain in the form of burning during urination, or pressure/discomfort felt on the urinary tract/pelvic floor [9]. A log function is available to record when the user experiences pain, the type of pain (i.e. burning during urination, or pressure), and the severity.
- **Menstruation log:** The app prototype is also tailored according to the user. For example, if the user identifies as being female, features specific to menstruation symptom fluctuations will be present. Particularly, research suggests that symptoms may flare during perimenstruation [4]. Therefore, the calendar feature will also track menstruation cycles to potentially plan for pain management.
- **Report generator:** The app prototype includes a report generator (Figure 2) whereby values entered into the log functions (i.e. diet, activity, sleep, and menstruation logs) will be compared to the pain log. This function will help to potentially identify what triggers pain. For example, the user may notice that they have more pain during times when they have consumed orange juice. Users can then discuss these correlations with their physician in order to identify flare triggers. A disclaimer will be included, alerting the user to always discuss any potential patterns they find with their physician before making any lifestyle changes to ensure appropriate use of the app.

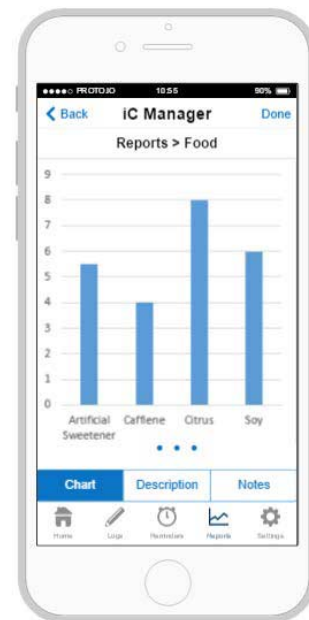


Figure 2– Report generator

- **Medication reminders:** The app prototype also has a feature where users can enter the medications that they are taking, and an alert will remind them when to take their next dose.
- **Appointment reminder:** An appointment reminder function is present to alert users of upcoming appointments with their family physician, urologist, or other specialist. As well, people with IC/PBS may seek chiropractic, biofeedback, psychotherapy,

acupressure/acupuncture, and massage therapy to manage their symptoms [4], further warranting an appointment reminder function.

- **Mental health and social resources:** A study by Rabin et al. [4] found that people living with IC/PBS reported more depression than the general population [4]. These authors note that people with IC/PBS could benefit from psychotherapeutic intervention, to increase one's sense of self-efficacy, and minimize self-stigmatization [4]. Research also notes how people living with IC/PBS may face challenges with respect to pursuing social, work, and recreational activities [4]. Thus, this app prototype includes a feature that provides resources for users to access, such as a list of professional counselors and physicians in one's area who specialize in depression in people living with chronic diseases.
- **Intimacy resources:** A previous study found that 88% of female participants with IC/PBS experienced sexual dysfunction [10]. Similarly, male adults with IC/PBS can experience pain with intercourse [11]. Thus, information will be available on treatment options such as physiotherapy, and cognitive behavioural therapy.
- **Public washroom locator:** A public washroom locator is also incorporated which identifies facilities in the user's vicinity (based on their current location), as urination frequency and urgency are common symptoms [2]. Many public washroom apps are already available; therefore, this app includes links to these resources.
- **Pregnancy:** People with IC/PBS who are pregnant may face challenges such as worsening of symptoms. As well, certain medications may be contraindicated in those who are pregnant [12]. Therefore, informational resources will be available to support users who are considering pregnancy or who are pregnant.

Based on available literature regarding the symptoms and challenges of living with IC/PBS, features were incorporated into an app prototype design for chronic disease management.

Discussion

As no other app is currently available for IC/PBS management, this work offers design features that may be helpful for those with the disease. Each feature was designed based on symptoms and challenges present in the available literature that people with IC/PBS face, demonstrating how research can be incorporated into the development of a health app.

This work represents the preliminary design of an app for IC/PBS. The researcher plans to undergo several iterations of evaluation, examining both the usability and content before developing the app further. The prototype will be of high fidelity, meaning that all functions (e.g. buttons, menus, input) will execute. Each possible screen will be evaluated using the Evidence-based Heuristics for eHealth Literacy and Usability [13]. The content will be evaluated by experts such as general practitioners and urologists. After several iterations of evaluation, representative users (i.e. people who have IC/PBS) will be asked to evaluate the app. Once the evaluation of the design is complete, the app will be ready to be developed in collaboration with professional developers.

Previous research has explored how social, political, and economic forces can influence the gendered construction of health information technology [14]. However, these researchers note a lack of examples of how to consider gender when developing health information technology [14]. Although this research will offer an example of making health information technology inclusive, it will also be challenging to navigate this space, in the absence of defined approaches.

Another challenge to designing this app will be to not perpetuate the discourse that the female body is risky and in need of surveillance [15]. The app is designed for users to manage their IC/PBS and learn what triggers their flares; however, many of the features described in this app are logs (i.e. diet, menstruation, activity, sleep, and pain logs). Perhaps this presentation is inappropriate, and another format should be considered. Further research is warranted to determine the best approach to IC/PBS management that does not reinforce oppression.

The search to find relevant apps was limited to English; therefore, it is possible that an IC/PBS management app exists that is available in other countries or in another language. Many general pain management apps are available, which people with the disease may be using. Similarly, people with IC/PBS may be using apps for individual symptoms such as medication reminders or a food/beverage diary.

Conclusion

Overall, this research outlines the design of an app that is inclusive to all users, based on known symptoms and challenges of living with IC/PBS. It contributes to the limited body of literature on incorporating sex and gender into health informatics research. Features of the app prototype include reminders, logs, and resources that could be helpful for people with IC/PBS to manage their condition and learn what triggers flares. The prototype will later undergo several iterations of evaluation for usability and content. Future work will involve developing this into a fully functioning app.

Acknowledgements

The author would like to acknowledge support from the Ontario Graduate Scholarship and her academic supervisors.

References

- [1] Hanno PM, Burks DA, Clemens JQ, Dmochowski RR, Erickson D, Fitzgerald MP, Forrest JB, Gordon B, Gray M, Mayer RD, Newman D. AUA guideline for the diagnosis and treatment of interstitial cystitis/bladder pain syndrome. *The Journal of Urology*. 2011 Jun 30;185(6):2162-70.
- [2] Berry SH, Elliott MN, Suttrop M, Bogart LM, Stoto MA, Eggers P, Nyberg L, Clemens JQ. Prevalence of symptoms of bladder pain syndrome/interstitial cystitis among adult females in the United States. *The Journal of Urology*. 2011 Aug 31;186(2):540-4.
- [3] Suskind AM, Berry SH, Ewing BA, Elliott MN, Suttrop MJ, Clemens JQ. The prevalence and overlap of interstitial cystitis/bladder pain syndrome and chronic prostatitis/chronic pelvic pain syndrome in men: results of the RAND Interstitial Cystitis Epidemiology male study. *The Journal of urology*. 2013 Jan 31;189(1):141-5.
- [4] Rabin C, O'Leary A, Neighbors C, Whitmore K. Pain and depression experienced by women with interstitial cystitis. *Women & health*. 2001 Apr 6;31(4):67-81.
- [5] Interstitial Cystitis Association. IC flares. [updated 2015], [Cited Dec 15, 2016], Available from: <http://www.ichelp.org/about-ic/symptoms-of-ic/icflares/>

- [6] Bassaly R, Downes K, Hart S. Dietary consumption triggers in interstitial cystitis/bladder pain syndrome patients. *Female pelvic medicine & reconstructive surgery*. 2011 Jan 1;17(1):36-9.
- [7] iTunes. Interstitial cystitis network food list. [updated 2016], [Cited Nov 18, 2016], Available from: <https://itunes.apple.com/ca/app/icn-food-list/id489794088?mt=8>
- [8] Goin JE, Olaleye D, Peters KM, Steinert B, Habicht K, Wynant G. Psychometric analysis of the University of Wisconsin Interstitial Cystitis Scale: implications for use in randomized clinical trials. *The Journal of urology*. 1998 Mar 31;159(3):1085-90.
- [9] Bogart LM, Berry SH, Clemens JQ. Symptoms of interstitial cystitis, painful bladder syndrome and similar diseases in women: a systematic review. *The Journal of urology*. 2007 Feb 28;177(2):450-6.
- [10] Bogart LM, Suttorp MJ, Elliott MN, Clemens JQ, Berry SH. Prevalence and correlates of sexual dysfunction among women with bladder pain syndrome/interstitial cystitis. *Urology*. 2011 Mar 31;77(3):576-80.
- [11] Interstitial Cystitis Association. Intimacy and IC. [updated 2015], [Cited Dec 17, 2016], Available from: <http://www.ichelp.org/about-ic/women/icandintimacy/>
- [12] Interstitial Cystitis Association. Pregnancy and IC. [updated 2015], [Cited Dec 17, 2016], Available from: <http://www.ichelp.org/about-ic/women/pregnancy-ic/>
- [13] Monkman H, Griffith J, Kushniruk A. Evidence-based heuristics for evaluating demands on ehealth literacy and usability in a mobile consumer health application. *Stud Health Technol Inform*. 2015 Aug 12;216:358-62.
- [14] Armstrong P, Armstrong H, Messing K. Gendering work? Women and technologies in health care. In *Gender, Health and Information Technology in Context 2009* (pp. 122-137). Palgrave Macmillan UK.
- [15] Green E, Griffiths F, Lindenmeyer A. 9 'It Can See into Your Body': Gender, ICTs and Decision Making about Midlife Women's Health. In *Gender, Health and Information Technology in Context 2009* (pp. 157-176). Palgrave Macmillan UK.

Address for correspondence

Corresponding author: Janessa Griffith

email: janessa.griffith@mail.utoronto.ca

cHRV Uncovering Daily Stress Dynamics Using Bio-Signal from Consumer Wearables

Tian Hao^a, Henry Chang^a, Marion Ball^a, Kun Lin^a, Xinxin Zhu^a

^a Center for Computational Health, IBM T.J. Watson Research Center, Yorktown Heights, NY, U.S.A.

Abstract

Knowing the dynamics of one's daily stress is essential to effective stress management in the context of smart and connected health. However, there lacks a practical and unobtrusive means to obtain real-time and longitudinal stress information. In this paper, we attempt to derive a convenient HRV-based (heart rate variability) biomarker named cHRV, which can be used to reliably reflect stress dynamics. cHRV's key advantage lies in its low maintenance and high practicality. It can be efficiently calculated only using data from photoplethysmography (PPG) sensors, the mainstream heart rate sensor embedded in most of the consumer wearables like Apple Watch. Benefiting from the proliferation of wearables, cHRV is ideal for day-to-day stress monitoring. To evaluate its feasibility and performance, we have conducted 14 in-lab controlled experiments. The result shows that the proposed cHRV has strong correlation with the stress dynamics ($r > 0.95$), therefore exhibits great potential for continuous stress assessment.

Keywords:

Stress, Psychological; Biomarkers; Heart Rate

Introduction

Stress is a feeling of psychological and physical tension in reaction to a challenge or demand. We all have experienced stress from time to time, if not every day. In a nutshell, stress is hard-wired as an "alarm system." When our brain perceives threat, it signals our body to release stress hormones to prepare for the "fight-or-flight" response. Once the stressor is gone, the alarm system is meant to be reset so that our body can recover to a normal and relaxed state. Unfortunately, in modern life, the nonstop and pervasive stressors tend to always keep us on high alert, which over time, could lead to a broad range of health problems ranging from headache, depression to heart diseases [1-3]. This is why stress management is especially important more than ever, and it's needed everywhere, especially places with high concentration of stress such as workplace, classroom etc.

According to a national survey conducted by American Psychological Association in 2012 [4], Americans consistently report stress levels that exceed what they believe is healthy. Specifically, in 2012, approximately seven in ten Americans reported that they experienced physical or non-physical symptoms of stress, including irritability or anger, fatigue and changes in sleeping habits. What is more troubling is that the survey unveils the fact that people are struggling to manage their stress and tend to choose ineffective activities as their

coping mechanism. For example, sixty-two percent of adults report that the activities they use to manage stress involve prolonged screen time, such as browsing the Internet, watching TV, playing video games and etc.

Despite the growing evidence of stress's epidemic impact on health, there still lacks practical solutions that offer support and proper intervention that help people navigate through their daily stress. This is largely due to the fact that there isn't a means to obtain people's daily stress dynamics in an unobtrusive manner. Measuring the levels of stress hormones, such as cortisol, is considered to be the "gold standard" and is able to provide an objective measure of stress level. However, it involves sampling and testing the subject's saliva or blood, which is invasive and time-consuming. Although providing accurate measurement, the result is only able to reflect only a snapshot rather than the dynamics of one's stress. Therefore, methods involving hormone testing are ill-suited for stress management.

In this paper, we propose a convenient HRV-based biomarker, (cHRV), that can be calculated only using photoplethysmography (PPG) data available in most consumer wearables. Due to its low maintenance and high practicality, cHRV bears great potential to make automated and personalized stress management possible. Not relying on custom devices other than a smartwatch, the computation of cHRV is designed to be passive and transparent to users with no extra cost, enabling a seamless integration into users' daily life. Moreover, cHRV offers a continuous and real-time stress signal, which is not only valuable at the individual level, but also powerful in generating insights about stress when examined at a higher level, for example, workplace, classroom and etc.

Specifically, in order to reliably reflect one's stress dynamics, cHRV is calculated using a number of physiological features based on heart rate variability (HRV), which is a commonly used indicator of Autonomic Nervous System (ANS) activities [5]. Studies show people under mental stress demonstrate a decrease in HF (high frequency) of HRV compared to a control group. Moreover, a sizeable body of research has also been dedicated to studying the link between HRV measurement and level of stress. The results suggest that HRV is a strong discriminative feature for distinguishing between stress and non-stress [6-8]. The changes in HRV is linked to the occurrence of stressors and is linked to each other due to the fact that the cardiovascular system is mostly regulated by the ANS through sympathetic and parasympathetic activities, which are also responsible for controlling body's reaction to subjective stressors. Therefore, HRV-based signals can offer insights into the activity of sympathetic and parasympathetic

pathways, which in turn can reflect physiological stress in certain contexts.

The contributions of this paper are as follows:

- We propose a method of extracting a HRV-based biomarker (cHRV) from PPG that is reflective of the stress dynamics. The method is convenient and practical, and has great potential to offer support and proper intervention that help people navigate through their daily stress.
- We present the preliminary evaluation results that suggest the proposed biomarker is highly correlated with the stress dynamics.

Methods

The data collection was conducted in a controlled, in-lab setting. The primary goal of the experiment was to investigate the feasibility and effectiveness of using cHRV to capture stress dynamics. Therefore, all the experiments shared a fixed structure (shown in Figure 1) which was pre-defined to isolate and control the stressor that causes the stress.

Controlled In-lab Experiment



Figure 1 – Structure of controlled in-lab experiment

Setting

The experiments were performed in either office rooms or reserved meeting rooms where only one or two researchers and one subject were present. Prior to the experiment, the researchers will assist the subject with putting on the devices for data collection, and make sure data is being recorded properly. The devices used for recording physiological data include; 1) a wrist-worn device with PPG sensor and electrodermal activity sensor (EDA), 2) an electrocardiogram (ECG) sensor and 3) a headset with a 4-lead electroencephalogram (EEG) sensor.

Procedure

As shown in Figure 1, the in-lab experiment consists of 3 sessions including Baseline, Stress Test and Recovery. Specifically, during the Baseline session, we play guided meditation for 20 minutes to relax the subject as much as possible, in an attempt to minimize the residual stress from other prior stressors, if there is any. Therefore, the collected physiological signs at the end of the Baseline session should reflect the subject's baseline state (when not under stress).

In the Stress Test session, we conduct a standard stress-inducing test where the subject is requested to answer a verbally asked, non-trivial arithmetic question (e.g., $2010-37=?$) every 10 seconds for about 6 mins [9-10]. This session consists of two such math tests with a 5-minute relax in between them to protect the subject from being under excessive stress. With the stress test, we exposed the subject to two types of typical daily stressors, which are 1) the stress as a result of being requested to solve non-trivial problems, and 2) the stress from having to finish tasks under time pressure.

In the Recovery session, we use the same relaxation technique as used in the Baseline session to help the subject recover from possible elevated stress. Data collected during this session will be used to investigate the recovery process from a stress buildup.

At the end of each session, we asked the subject to rate his or her current perceived stress level (PSL) on a scale from 0 to 10, with 0 being not stressed at all and 10 being extremely stressed.

Table 1 – List of Data Collection

Data	Sensor	Sampling Rate
PPG signal	PPG	64Hz
Skin Conductance	EDA	4 Hz
Brain Wave	EEG	250 Hz
Heart Rate	PPG	1 Hz
Acceleration	ACC	32 Hz
Perceived Stress Level	Self-report	n/a

Subjects

The subjects consist of 12 IBM employees, mostly composed of young adults with a few middle-aged subjects. Our study along with its data collection procedure is approved by the Institutional Review Board. All the subjects voluntarily agreed to contribute to the data collection and signed a consent form. Individuals will be excluded if they have significant health conditions or take medications that interfere with stress tasks including diagnosed cardiovascular conditions (e.g., arrhythmia, hypertension), neurological disorders (e.g., seizure disorder, stroke, transient ischemic attack), mental illness (e.g., depression, panic disorder) and cognitive or attention disorders (e.g., attention deficit/hyperactivity disorder).

Dataset

During the course of the experiment, we have collected a rich set of raw data from various sensors (shown in Table 1). For performance evaluation, we extracted cHRV along with three other signals commonly used as indicators of stress for comparison (shown in Figure 3).

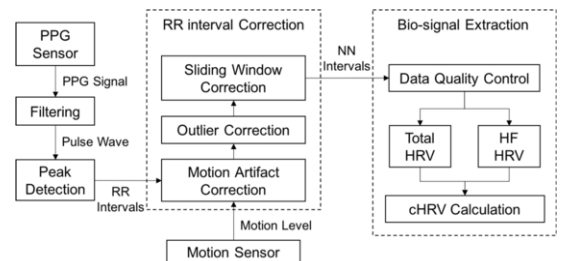


Figure 2 – A flowchart illustrating the process of extracting cHRV from consumer wearables

Extracting cHRV from PPG Signal

PPG sensor has been widely embedded in most smartwatches and fitness trackers. Similar to pulse oximeter, PPG is a light-based technology to sense the rate of blood flow as controlled by the heart's pumping action. As shown in Figure 2, A series of operations are involved in the process of extracting cHRV from consumer wearables with motion sensor and PPG sensor.

To calculate RR intervals from PPG signal, we first perform a 6-order Butterworth low-pass filter on the original PPG signal, with a cutoff frequency of 2 Hz. This filtering process is intended to filter out noise irrelevant to heart beats as much as possible. Then a peak detection operation will be conducted on the filtered PPG signal to identify peaks in the signal that represents heart beats. Lastly, the time intervals between successive identified peaks are extracted, resulting in a temporal sequence of RR intervals.

A sequence of outlier-free, beat-to-beat intervals is essential to accurate HRV calculation. In order to obtain normal and reliable RR intervals, commonly referred to as NN intervals, we filter out the outliers in RR intervals with a correction procedure consisting of three steps. First, as motion artifact is the major cause to corrupted PPG signal, we filter out the RR intervals based on their corresponding motion level, which is derived using acceleration data collected from the motion sensor. In the second step, a standard threshold-based method is used to filter out abnormal RR intervals based on its duration and also the difference in duration between consecutive RR intervals. Lastly, a distribution-based method is used to identify and remove outliers within a sliding window.

After the RR interval correction, the resulted sequence of the NN intervals will be used to extract cHRV. In clinical practice, no less than five minutes of NN intervals is required to calculate short-term HRV features, although recent research studies have suggested that shorter windows (e.g., 60 seconds) [11] may also be sufficient, especially for time-domain methods. In our case, a five minute sliding window of NN intervals are used to extract cHRV. To ensure the reliability of the feature extraction, a Data Quality Control component is responsible for identifying the window with a high percentage of discarded RR intervals, which is a major sign of poor data quality, mostly caused by excessive motion. Only data within windows, with acceptable data quality, are used to calculate HRV-based features. Features including the total power of HRV SDNN (standard deviation of normal to normal R-R intervals) and the high frequency power rMMSD (square root of the mean squared difference of successive N-N intervals) of HRV will be combined to calculate cHRV.

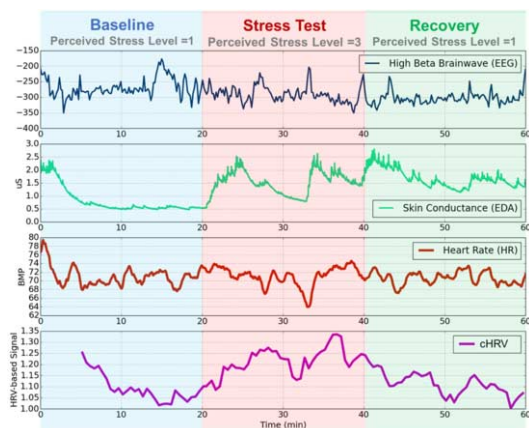


Figure 3 – Stress signals extracted from a 3-session in-lab experiments (from top: EEG, EDA, HR and cHRV signals), along with self-reported stress level at the end of each session.

Other Stress Signals Extracted for Comparison

To gauge the performance of cHRV in reflecting stress dynamics, we extract three other signals that are commonly used as indicators of stress and compare them with cHRV for evaluation purpose. Data used to extract these signals were all collected in parallel with the data used to derive cHRV. These stress signals include:

- **Skin Conductance (EDA):** The increase in EDA (Electrodermal Activity) signal indicates increased sweat production, which is commonly associated with sympathetic arousal.
- **Power of high-beta band of the brain wave (EEG):** increased power is associated with high arousal (e.g., stress and alertness)
- **Heart Rate (HR):** Certain stressful situations could lead to increased heart rate.

In Figure 3 we plot the cHRV signal along with the EDA, EEG and HR signals over time using data collected from a typical in-lab experiment. In these experiments, the subject's stress level has been successfully elevated by the Stress Test session and later relieved by the Recovery session, resulting in an increase of perceived stress level from 1 to 3 followed by a decline from 3 to 1. A key observation of this case is that the EDA and cHRV are the signals that most reflected the subject's stress dynamics. Specifically, a noticeable decline can be observed in both signals during the Baseline and Recovery sessions, which are designed to reduce previous stress and induced stress, respectively. More importantly, the increase in stress during the Stress Test session is clearly reflected in both EDA and cHRV. Interestingly, the effect of the short five minute resting in the middle of the Stress Test session is also captured, resulting in a brief dip around the thirtieth minute in both EDA and cHRV signals.

Results

The primary goal of the evaluation is to examine the cHRV's ability and effectiveness in reflecting the dynamics of stress. The basic idea is to examine the correlation between a signal and the stress dynamics during the experiment.

Perceived Stress Level

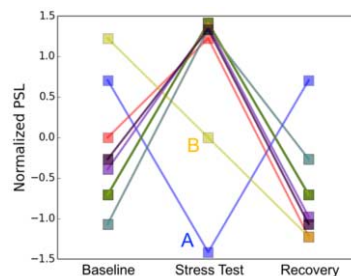


Figure 4 – Subjects' Perceived Stress Levels (PSL) during the in-lab experiment sessions (plotted with data from 14 experiments)

For evaluation purposes, we use Perceived Stress Level (PSL) as the proxy of the stress dynamics. PSL is reported by the subject at the end of each session on a scale from 0 to 10. The structure of our in-lab experiment was designed to see if the experiment showed a result in an PSL increase during the Stress Test session, and a PSL reduction after the Recovery

session. In Figure 4, the normalized PSLs collected from 14 experiments are plotted over three sessions. Note that only two subjects' PSL reports are inconsistent with the structure of the experiment. Interestingly, one subject (A) reported reduced PSL in the Stress Test session (blue line Figure 4), although, all four stress signals suggest otherwise (blue lines Figure 5). Therefore we speculate that subject (A) might have a misperception of his or her stress level. Another subject (B) reported constant stress reduction (yellow line Figure 4) during the experiment. This could be largely because the subject began the experiment with relatively high residual stress built up from work or other previous activities, rendering the stress-inducing session ineffective. Although the Stress Test did not work as intended, the cHRV successfully captured data that showed a constant stress reduction over the experiment (the yellow line shown in Figure 5d).

Correlation-based evaluation

To investigate and demonstrate cHRV's effectiveness in reflecting stress dynamics, we calculate all four aforementioned stress signals' correlation with the subjects' perceived stress level for each of the experiments.

Specifically, we first conduct simple processing on the signal contained in each session, resulting in a sequence of three values for each signal. Due to inherent nature of the three sessions, the value is calculated by averaging the last five minutes of signals for Baseline and Recovery sessions, and the entire twenty minutes of signals for Stress Test session. Figure 5 shows the normalized result for each signal. Lines with the same color in Figure 4 and 5 represent data from the same experiment.

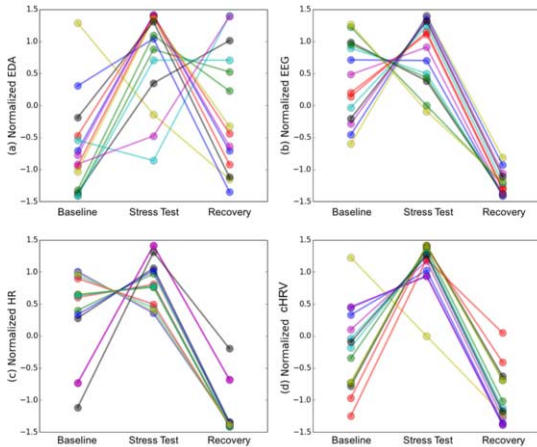


Figure 5 – Comparing cHRV with other signals in reflecting stress dynamics of 14 in-lab experiment.

Next, for each experiment, we calculate Pearson Correlation between the corresponding stress signal shown in Figure 5 and the reported PSL shown in Figure 4. The result is plotted in Figure 6. We can see that cHRV achieves the highest average correlation ($r > 0.95$) with the most reliable performance. In contrast, other signals such as EDA and EEG failed in reflecting the stress dynamics, resulting in lower correlation with PSL. As we can see in Figure 5a, although EDA signal is able to capture the stress dynamics in most of the experiments, there are several results where the elevated stress in the Stress Test session were incorrectly measured as reduced stress. This could be mainly because that the skin conductance is also

affected by other environmental factors such as room temperature, therefore adding uncertain noises. In Figure 5b, we can see that the EEG signal is accurate in detecting the decline in stress level in Recovery session, but tends to yield unreliable measurement for Baseline session.

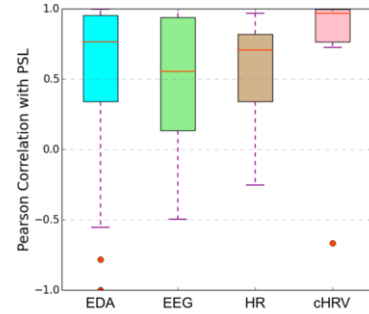


Figure 6 – Pearson Correlation (90 percentile) between 4 stress signals and the corresponding Perceived Stress Level.

Discussion

In this experiment, we use perceived stress level as the proxy stress indicator because there lacks ground truth on exactly how stressed people are. However, we have learned from the subjects that there may exist a gap between one's psychological perception of stress feeling and physiological measurement of stress response, as illustrated by Figure 4 that subject (A) might have a misperception of his or her stress level. To further probe this problem, we propose to categorize the perception-measurement levels into four groups, as shown below:



Figure 7 – Relationship between perceived stress level and measured stress level.

However, the current experiment is limited by the number and diversity of subjects. In the next phase of this study, we will recruit more subjects, then categorize them according to these four groups, and analyze for each group what characteristics are representative and what factors or context contribute to the misperception (Groups 1 and 4). This line of research work will help us establish unique user stress profiles and identify influencing stressors and contexts. The goal is to provide in time, continuous feedback to users so they have a better self understanding on how their minds and bodies function and respond to various stressors.

This is an essential step towards stress management. Our proposed cHRV method also offers a valuable objective addition to widely used psychological instruments for measuring self-reported perceived stress scales (PSS) [12] by providing real-time convenient physiological measures in daily life.

Another valuable insight gained from this experiment is that same physiological features could mean different stress indexes for different users, as shown in Figures 5 and 6. Therefore, one-size-fits-all detection model will not provide accurate and meaningful results for everyone. The proposed stress monitoring approach takes into account the individualized stress profile, which will be adjusted, using user's sparse stress labeling (e.g., users' perceived stress level at a certain time), so that it can gradually adapt to the user's unique physiological response.

Driven by the proliferation of wearable devices, the authors believe the ability to continuously monitor and manage stress in real-time is a critical component of this new exciting commercial domain. Although future studies are needed, the initial results of cHRV from PPG sensors, the mainstream heart rate sensor embedded in most of the consumer wearables, has shown its potential to not only enable researchers to explore practical, continuous, unobtrusive and personalized stress management, but also empower users to stay aware of their stress level in real-time and effectively and efficiently manage their stress on a daily basis.

Conclusion

In this study, we experiment practical and unobtrusive means to obtain real-time and longitudinal information about stress. The proposed cHRV approach uses proliferated consumer wearables to derive a convenient HRV-based biomarker to reflect daily stress dynamics. We compare and evaluate the feasibility and performance of cHRV through in-lab controlled experiments with other biosensors, including EEG, EDA and HR. The result shows that the proposed cHRV has strong correlation with the stress dynamic, and therefore exhibits great potential for continuous daily stress assessment with reasonable reliability and high practicality.

Acknowledgements

This experiment uses Empatica E4 wristband [13].

References

- [1] G. P. Chrousos and P. W. Gold, "The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis," *Jama* **267** (1992) no. 9, 1244–1252.
- [2] N. Schneiderman, G. Ironson, and S. D. Siegel, "STRESS AND HEALTH: Psychological, Behavioral, and Biological Determinants," *Annu. Rev. Clin. Psychol.* **1** (2005) 607–628.
- [3] M. Al'Absi and D. K. Arnett, "Adrenocortical responses to psychological stress and risk for hypertension," *Biomed. Pharmacother.* **54** (2000) no. 5, 234–244.
- [4] American Psychological Association, "Stress in America: Missing the Health Care Connection. <https://www.apa.org/news/press/releases/stress/2012/full-report.pdf>," 2012.
- [5] G. G. Berntson et al., "Heart rate variability: origins, methods, and interpretive caveats," *Psychophysiology* **34** (1997), no. 6, 623–648.
- [6] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (2014), 2957–2960.
- [7] J. F. Thayer, F. Ahs, M. Fredrikson, J. J. Sollers, and T. D. Wager, "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neurosci. Biobehav. Rev.* **36** (2012), no. 2, 747–756.
- [8] T. G. Vrijkotte, L. J. Van Doornen, and E. J. De Geus, "Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability," *Hypertension* **35** (2000), no. 4, 880–886.
- [9] Z. B. Moses, L. J. Luecken, and J. C. Eason, "Measuring task-related changes in heart rate variability," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, (2007), 644–647.
- [10] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *Eur. J. Appl. Physiol.* **92** (2004), no. 1–2, 84–89.
- [11] M. R. Esco, & A.A. Flatt, Ultra-Short-Term Heart Rate Variability Indexes at Rest and Post-Exercise in Athletes: Evaluating the Agreement with Accepted Recommendations. *Journal of sports science & medicine*, **13**(3) (2014),535.
- [12] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *J. Health Soc. Behav.*, (1983) 385–396..
- [13] Empatica. <https://www.empatica.com/>.

Address for correspondence

Tian Hao. E-mail: thao@us.ibm.com

Who Is Your Doctor? Analysis of Patient-Reported and EHR-Imputed Primary Care Physician

Shan He, PhD^a, Sue Dintelman, MS^b, Jake Sangster^a,
Darren K. Mann^a, Thierry Guillermin, PhD^a, Sidney N. Thornton, PhD^{a,c}

^a Homer Warner Center for Informatics Research, Intermountain Healthcare, Salt Lake City, Utah, USA

^b Utah Health Information Network, Murray, Utah, USA

^c Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA

Abstract

Significant efforts have been made to improve physician-to-physician communication and care coordination during transition of care in order to reduce adverse events and readmissions. As electronic health records (EHRs) become widely available, many hospitals have implemented physician collaboration and hand-off tools to automatically send admission notifications, discharge summaries, and pending laboratory results to a patient's primary care physician (PCP). However, the effectiveness of such tools depends on a fundamental question that remains unstudied: who is the patient's PCP? Missing or outdated PCP information may become the bottleneck to effective patient-centered care coordination regardless of existing efforts on promoting interoperability among healthcare providers. In this paper, we characterized patient-reported PCPs and experimented with an imputation algorithm that automatically infers a patient's primary provider based on patient-provider encounter data. We compared the imputation results with patient-reported PCPs and suggested practical uses of our findings.

Keywords:

Physicians, Primary Care; Patient-reported; Care Coordination

Introduction

Primary care physicians (PCPs, also known as general practitioners) play a critical role in providing continuing care of patients, especially those with complex health needs [1]. PCPs coordinate care among various care team members of a patient by communicating with hospitals, emergency departments, specialty care, home care and social service providers. Studies have reported that increased involvement of PCPs is associated with significant decreases in healthcare utilization [2]. In United States, diverse stakeholders, including Federal and State agencies, insurers and employers, are engaged in numerous efforts to promote the shift to new healthcare delivery models such as Patient-Centered Medical Homes (PCMH) [3] and Accountable Care Organizations (ACO) [4]. These models reward for value rather than volume and aim to improve clinical outcomes. Both the PCMH and ACO models rely on primary care physicians to coordinate care across the healthcare continuum. For example, patients discharged from the hospital are one major group being targeted for care coordination. Numerous studies have shown that timely outpatient follow-up is associated with lower readmission rates [5,6]. However, studies also reported that poor communication among physicians across care settings have posed practical challenges

to coordinated care. In a study of primary care physicians across the industrialized world, only 31% of physicians surveyed said that they were always notified when a patient is discharged from the hospital and 32% for emergency department visits in United States [7]. Similarly, communication between PCPs and specialists regarding referrals and consultations is often inadequate. A study reported that 80% of specialists said they "always" or "most of the time" send consultation results to the referring PCP, but only 62% of PCPs said they received such information [8].

To promote effective communication among physicians during transition of care, most previous studies focused on designing hand-off tools or evaluating the availability and content of discharge summaries [9,10]. As the adoption of EHRs among hospitals and individual practices increases, many EHRs are configured to automatically fax or electronically send discharge summaries to PCPs upon patient's discharge. However, such automated services rely on accurate and complete PCP information captured in EHR or associated administrative systems. In other words, the practical value of the transition of care tools depends on the assumption that the hospital or specialist knows who the patient's PCP is and where to send discharge summaries or consultation notes. In reality, the PCP is not reliably captured. Hospitals are often faced with the challenge of missing or outdated PCP information in their system. Discharge summaries may be sent to an invalid fax number or the wrong physician if the patient has changed PCP for various reasons. The inaccurate PCP information causes communication lapses and poses significant privacy concerns considering the protected health information (PHI) contained in the documents.

Research into the clinical implications of accurate patient-PCP relationships has deferred to the focus of measurement for the operational payment imperative. In the health insurer industry, in order to measure a provider's performance, members (patients) are attributed to a provider through an analysis of healthcare claims [11]. Attribution rules vary across organizations, depending on the objectives and application. However, this patient attribution approach only focuses on insurance members and providers who participate in certain incentive programs such as the ACO delivery model or a specific health plan network. Our study examines the PCP relationship from a patient's perspective and is targeted at the complete patient population, including those uninsured and underinsured.

In this study, we focused on answering some fundamental questions regarding patient PCPs. How many patients have a PCP? Do all patients know who their PCP is? Do all patients

know what “PCP” refers to? Is patient-reported PCP information reliable? What should we do when the PCP information is missing? The specific objectives of this study were to a) understand patient-reported PCP information, b) demonstrate the feasibility of using EHR data to impute a patient’s PCP, and c) compare the imputed PCP with patient-reported PCP.

Methods

Setting

Intermountain Healthcare (hereafter referred to as “Intermountain”) is a not-for-profit health system based in Salt Lake City, Utah that operates 22 hospitals, over 185 clinics, a broad range of laboratory and pharmacy services, and an affiliated health insurance company. Intermountain has about 1,400 employed physicians and 3,000 affiliated physicians. It is the largest integrated health system in the Intermountain West region of the United States.

Data Source

Data for this project included enterprise-wide inpatient, outpatient, and emergency department registration and encounter records at all Intermountain facilities.

Registration Dataset

To understand the characteristics of patient-reported PCP, we analyzed the registration data generated from two patient registration applications—one for hospital and emergency department encounters (R1) and one for outpatient clinic encounters (R2). R1 requires the registrar to collect encounter-based PCP information, which means that the patient needs to declare his/her PCP at each visit, regardless of whether the PCP information has been collected previously. R2 treats PCP as longitudinal data and allows the registrar to keep or update the PCP information if the patient has reported a PCP during previous visits. However, in R2, new PCP information simply overwrites the old information and current data do not support the analysis of a patient’s PCP history. R2 also allows the registrar to enter the reason for missing PCP. We extracted registration data from both R1 and R2 during a two-year period between September 1, 2014 and September 1, 2016.

Encounter Dataset

To examine the feasibility of using EHR data to impute a patient’s PCP, we analyzed patient encounter data to extract all patient-provider interactions during a five-year period between September 1, 2011 and September 1, 2016. The encounter dataset included all ambulatory visits at any Intermountain facility. Inpatient hospital visits and emergency department visits were excluded based on the assumption that hospitalists and emergency medical doctors focus on individual encounters and do not establish a longitudinal relationship with a patient. Each encounter record included in the analysis consisted of a unique patient identifier, the provider’s identifier for that visit, and time and location of the visit.

We linked the registration dataset and encounter dataset with our internal provider directory that includes both Intermountain-employed providers and independent practitioners who are affiliated with Intermountain. Many of the independent practitioners have privileges to admit patients to an Intermountain hospital. However, they usually have separate patient registration and EHR systems for outpatient visits. Each record in the provider directory included a provider’s full name, specialty, employment status, practice location, phone number,

fax number, etc. However, many records in the provider directory were identified with incomplete information.

External Dataset

Considering the fact that patients frequently visit providers from multiple healthcare organizations, we extended the imputation analysis to statewide encounter data outside of our health system. We extracted all clinical messages received by the state Health Information Exchange (HIE) from major hospitals, clinics, labs and insurers in Utah between 2011 and 2016. The types of clinical messages included admit, discharge, and transfer (HL7 ADT), laboratory, radiology and transcription results, and prescriptions (HL7 Pharmacy/Treatment Encoded Order Message—RDE). Each clinical message extracted for this study consisted of a unique HIE patient identifier, provider information in the Patient Visit Information (PV1) segment, and date of service. The HIE patient identifier was linked to the Intermountain patient identifier through the community master patient index.

Analysis

The first part of the study was a descriptive analysis to characterize patient-reported PCPs. We analyzed the completeness of PCP information declared by patients or family members. We also analyzed the consistency of PCP declared across encounters for the same patient using the registration data from R1. Reasons for missing PCPs were summarized.

The second part of the study was to impute the primary provider for a subset of patients who had inpatient or emergency department encounters between September 1, 2014 and September 1, 2016. The imputation algorithm leveraged all records extracted into the encounter dataset and calculated a closeness score for each patient-provider pair using Eq.1 was an updated algorithm from our previous work [12]. The underlying design principle of our imputation algorithm is the more frequent and more recent a patient was seen by a provider, the higher the closeness score between the patient and the provider. Most existing attribution rules employed by health insurers are also based on this principle. However, these rules only use the date for the most recent visit when there is a tie between multiple providers. We further refined the algorithm by taking into account the date of each previous visit. Each patient-provider encounter record was assigned with a score that was inversely correlated to the lapsed time between the current date and the date of the encounter. The “current date” (denoted as “date of the target visit” in Formula (1) refers to the time when a provider is imputed based on all previous encounters. The sum of all visit scores for a unique patient-provider pair was the total score for that patient-provider relationship. A threshold (closeness score ≥ 2) was defined to identify the providers who were considered as the primary provider. The parameters k and e in Eq.1 and the threshold were chosen based on iterative adjustments in order to make the algorithm clinically intuitive.

$$\text{closeness_score} = \sum_{i=1}^n e^{-k(T-t_i)} \quad (1)$$

where

n = total number of encounters;

$k=0.001$;

T = date of the target visit;

t_i = date of the i^{th} encounter

We compared the imputed primary providers with patient-reported PCPs for two subsets of patients: 1) patients who had more than one inpatient or emergency department encounters during the study period and declared the same PCP across

multiple encounters, and 2) patients who had inpatient or emergency department encounters during the study period and did not declare a PCP at the encounter. Step 1) was to understand the strength and weakness of the imputation algorithm. Step 2) was to demonstrate the potential value of the imputation algorithm in identifying primary providers to be notified for inpatient and emergency encounters when the PCP information is missing.

Results

There were 2.5 million registration records for about 830,000 patients generated from R1 between September 1, 2014 and September 1, 2016. R2 had about 1 million registration updates during the two-year period.

Completeness of Patient-reported PCP

Table 1 shows the completeness of patient-reported PCP information by encounter type. About 58% of all inpatient encounters had some PCP documented during registration and 52% for emergency visits. Inpatient Behavioral health encounters had the lowest PCP declaration rate (24%). Outpatient registrations had the highest PCP declaration rate in general (74-89%). For patients who reported PCPs, the PCP information was not consistent across encounters. Among all inpatient and emergency encounters generated in R1, 31% of the patients had more than one encounter during the two-year period, only 27% of whom reported the same PCP across multiple encounters. Over 5,000 patients had more than five PCPs reported and the maximum number of PCPs reported by a patient (or family member) was 15 during the two-year period.

Table 1 – Patient-reported PCP by Encounter Type

Encounter Type	Proportion of Patient who Reported a PCP
Inpatient	58%
Maternity/newborn	42%
Behavioral Health	24%
Inpatient Transplant	75%
Inpatient Other	65%
Outpatient	N/A*
Laboratory	83%
Imaging	75%
Same-day Surgery	74%
Clinic	89%
Emergency	52%

*The outpatient registration records were extracted from a combination of R1 and R2. The overall proportion of outpatient PCPs does not apply here.

For patients with missing PCP during registration, only 0.2% had a reason documented. About a third of these patients reported that they did not have a doctor they see regularly, and 5% of the patients did not remember or only remember partial information about their PCP (Figure 1). Very few patients did not want to disclose their PCP information and did not want the hospital to forward the discharge summaries to their PCP.

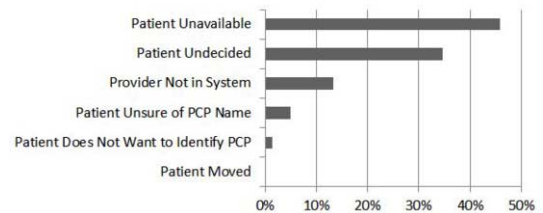


Figure 1 – Reason for Missing PCP at Registration

Provider Profiles

The top ten primary specialties of patient-reported PCPs are illustrated by the dark grey bars in Figure 2. Eleven percent of patient-declared PCPs were Family Medicine doctors, followed by Physician Assistants (6%). Surprisingly, Emergency Medicine doctors and Dentists were at the top ten most popular PCP specialties reported by patients. Due to the data quality issue of the provider database, not all providers have a primary specialty recorded. One percent of patient-reported PCPs were non-person accounts (e.g. facilities or clinics). Forty percent had an inactive status at the time of query. Eighty-one percent had either a phone number or fax number recorded in the provider directory. However, our analysis did not validate the phone number and fax numbers.

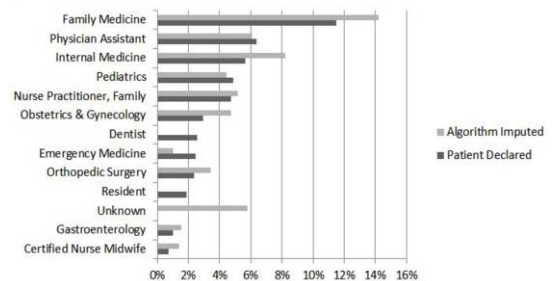


Figure 2 – Patient-reported vs. Imputed PCP by Specialty

As shown by the dark grey bars in Figure 3, about half of the patient-reported providers were independent practitioners, most of whom have their own practices and EHR systems. Some of the independent practitioners were affiliated with Intermountain hospitals and had access to the EHR while others do not.

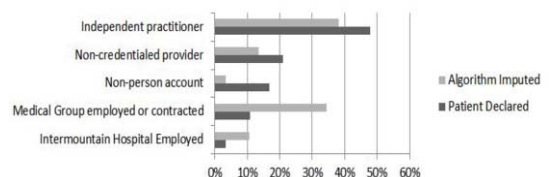


Figure 3 – Patient-reported vs. Imputed PCP by Employment Status

Algorithm-imputed PCPs

Firstly, we conducted the imputation algorithm on patients who reported the same PCP across multiple encounters during the

two-year period. The results showed that the algorithm identified one or more primary providers (closeness score ≥ 2) for 62% of the patients and 38% of the imputed providers were consistent with patient-reported PCPs. Table 2 shows some example closeness scores calculated by the imputation algorithm using the “Target Visit Date” as the reference point of time. Any visit occurred before the target visit date was counted in the closeness score. As shown by the highlighted rows for pt001, even though the number of visits with pv102 exceeded the number of visits with pv101, the closeness score for pv101 was higher since the visits were more recent.

For those patients who did not get an imputed provider, we manually reviewed some randomly selected patients’ chart. As expected, the primary factor was that these patients only had emergency visits at Intermountain. Reported PCPs from these patients were independent practitioners who were not employed by Intermountain. As a result, the visit history was captured by the doctor’s own EHR that was not part of the analysis dataset.

Table 2 – Example Closeness Scores
Calculated from Previous Visits

Patient ID	Imputed Provider ID	Target Visit Date	Total No. of Visits	Last Seen	Score
pt000	pv100	10/09/2015	5	08/26/2013	1.4
pt001	pv101	11/29/2015	7	08/03/2015	5.4
pt001	pv102	11/29/2015	17	11/14/2010	2.2
pt002	pv100	07/30/2016	38	07/08/2016	13.8
pt002	pv103	07/30/2016	10	01/22/2015	4.4
pt002	pv105	07/30/2016	3	04/28/2016	2.6

Note: all identifiers and dates were de-identified without affecting the trend of the algorithm.

Secondly, we applied the algorithm to patients who did not have a PCP declared at the inpatient or emergency encounter during the study period. The purpose of the second part of the imputation analysis was to demonstrate the feasibility of identifying the appropriate provider who should be notified about a hospital stay or emergency visit when the PCP was not collected from the patient. Results showed that the algorithm could identify one or more primary provider for 38% of the encounters.

We further compared the primary specialty and employment status of the algorithm-imputed providers with the patient-reported PCPs. As shown in Figure 2, the top 5 specialties for imputed providers aligned with patient-reported PCPs. Dentist, resident and emergency medicine did not have or had very low representation in imputed providers. Figure 3 shows that the imputed providers were more likely to be Intermountain employed. This can attributed to the input dataset used by the algorithm.

Imputation Results from External Data

Considering the limitations of Intermountain-only encounter data, we conducted a preliminary analysis on a subset of patients (2,600) who had an inpatient stay at one of the Intermountain facilities between June 2015 and February 2016 and did not specify a PCP at registration. We queried the state HIE database to extract all clinical messages that indicated a patient-provider encounter. Results showed that 71% of the patients can be matched in the HIE dataset. A total number of

26,888 messages were received from 37 healthcare organizations on 693 (27%) patients. The distribution of message types is illustrated by Figure 4.

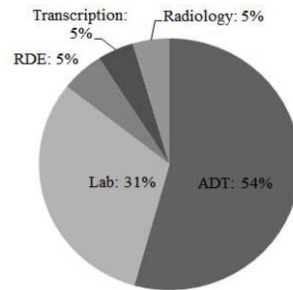


Figure 4 – Clinical messages received by the state HIE on sample patients

By applying the same imputation algorithm from Formula 1, we identified the primary provider for an additional 15% of the patients. As shown in Table 3, the completeness of the provider information contained in the PV1 segment of the HL7 messages was between 1% and 17%.

Table 3 – Completeness of Provider Information in
Clinical Messages(HL7) Sent to State HIE

Provider Segment	Proportion of Patients with Provider Information
PCP	9%
Referring Provider	10%
Ordering Provider	17%
Prescribing Provider	1%

Discussion

Our study suggest that PCP information is not well-documented in EHRs. Only about half of the inpatient and emergency encounters have PCPs reported by the patient or family members. For certain types of encounters such as behavioral health inpatient stays, the completeness of PCP information can be as low as 24%. Missing, incomplete or outdated PCP information can become the bottleneck to effective physician communication and care coordination, especially during transition of care. Based on our analysis, the reasons for missing PCPs were multifaceted. There may not be sufficient time to collect the PCP information during registration. Some patients may only remember their PCP’s last name or practice clinic.

For patients who reported a PCP during registration, it is not clear how reliable the PCP information is. As shown by our provider profile analysis on patient-reported PCPs, some patients declared an emergency doctor or dentist as their PCP. This may indicate that patients or family members do not understand what a PCP refers to. As part of the care coordination and outreach effort, Intermountain has been instructing designated staff to collect PCP information by visiting patient rooms at an inpatient facility if the PCP information was not documented during registration. We found that sometimes the hospital staff needs to ask the question in several different ways to explain what a PCP is. For example:

Which doctor do you usually see for your annual visit?

Who ordered your medications?

Who do you want us to send your discharge information and lab results to after you go home?

Our imputation analysis demonstrates a promising approach to infer a patient's primary provider based on encounter history. This will greatly benefit patients discharged from hospitals or emergency departments whose PCP was not captured during registration. The imputation results only showed 38% consistency with patient-declared PCP. There are several reasons for this: 1) Patient-reported PCPs are not the gold standard. It only reflects patients' perceptions, which may not be reliable as indicated by various findings in this study. In other words, the algorithm-derived primary providers may be more accurate since they are based on the actual patient-provider interactions; 2) The performance of the imputation algorithm depends on the comprehensiveness of the patient-provider encounter dataset as input. We only used enterprise wide encounter data for the first part of the analysis to infer the primary providers for 62% of the patients. As we extended the input to include state wide data, it increased the proportion by 15%.

Our study also suggests that the lack of a statewide or national provider directory is a limitation preventing both PCP information collection from patients and PCP imputation by the algorithm. A complete and up-to-date provider directory will facilitate provider lookup when the patient could not remember his/her PCP's full name, especially when the provider is outside of the health system. Most healthcare organizations maintain their own provider directory, which normally has up-to-date information for internally employed providers. However, as the adoption of interoperable health information technology such as Direct messaging [13] increases, the need to maintain providers from external organizations and their communication preferences grows. Currently, many hospital units maintain a separate list of commonly interacted providers (e.g. referral doctors) with phone number, address and fax number in a spreadsheet file. Some healthcare organizations exchange provider list that contains Direct email addresses with their partner organizations on a regular base. A formal provider directory management infrastructure could help to overcome the duplicate and uncoordinated efforts required from individual organizations.

Limitations and Future Work

The major limitation of our study is the lack of external encounter data as input to the imputation algorithm to infer PCPs outside of Intermountain. Although we conducted a preliminary analysis using the state HIE clinical messages received from other healthcare providers, the missing values in the provider information segment of the HL7 messages limited the performance of the algorithm. In addition, for those HL7 messages with provider information, there is no unique provider identifier across organizations. Each organization used its own provider identifier and some HL7 messages only had provider's name without an identifier.

Nevertheless, our analysis demonstrated feasibility of using EHR data to infer a patient's PCP. It is important to note that algorithmic imputation should not overwrite patient's preferences. The purpose of the imputation algorithm is not to replace patient-reported PCPs but to serve as a complementary mechanism to support accurate PCP information collection. As future work, we plan to display the imputed primary provider(s) in the registration systems and patient portals to allow patients to confirm or modify the list and specify their information disclosure preferences. PCP is longitudinal information and should be maintained across encounters regardless of the care setting. We plan to test the algorithm with other healthcare organizations and design a community care team framework to

allow organizations to share the imputed provider information through an ongoing statewide care coordination effort [14].

Conclusion

Effective and seamless communication with primary care physicians during the discharge period is pivotal to boosting patient safety and reducing the likelihood of avoidable readmissions. Our findings suggest that complete and reliable PCP information may not be available in EHRs and associated administrative systems. Lack of PCP information may hinder collaboration among physicians and delay timely follow-up. We demonstrated the feasibility of using historical encounter data that is easily accessible in any EHR to infer complementary provider information when a self-reported PCP is missing or invalid. Our methods can be generalized to any healthcare organization to improve the availability and accuracy of PCP information for care coordination purpose.

References

- [1] K. Davis, S.C. Schoenbaum, A.-M. Audet. A 2020 vision of patient-centered primary care. *J Gen Intern Med*. Springer; 2005 Oct;20(10):953–7.
- [2] S.J. Kravet, A.D. Shore, R. Miller, G.B. Green, K. Kolodner, S.M. Wright. Health Care Utilization and the Proportion of Primary Care Physicians. *Am J Med*. 2008 Feb;121(2):142–8.
- [3] Patient-Centered Medical Home PCMH [Internet]. [cited 2016 Dec 11]. Available at: <http://www.ncqa.org/programs/recognition/practices/patient-centered-medical-home-pcmh>
- [4] Accountable Care Organizations (ACO) [Internet]. 2015 [cited 2016 Dec 10]. Available at: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/index.html?redirect=aco/>
- [5] B.S. Brooke, D.H. Stone, J.L. Cronenwett, B. Nolan, R.R. DeMartino, T.A. MacKenzie, et al. Early primary care provider follow-up and readmission after high-risk surgery. *JAMA Surg*. NIH Public Access; 2014 Aug;149(8):821–8.
- [6] T.S. Field, J. Ogarek, L. Garber, G. Reed, J.H. Gurwitz. Association of early post-discharge follow-up by a primary care physician and 30-day rehospitalization among older adults. *J Gen Intern Med*. Springer; 2015 May;30(5):565–71.
- [7] R. Osborn, D. Moulds, E.C. Schneider, M.M. Doty, D. Squires, D.O. Sarnak. Primary Care Physicians In Ten Countries Report Challenges Caring For Patients With Complex Health Needs. *Health Aff (Millwood)*. 2015 Dec;34(12):2104–12.
- [8] A.S. O'Malley, J.D. Reschovsky, S. B, S. B, S. C.J, W. PT, et al. Referral and Consultation Communication Between Primary Care and Specialist Physicians. *Arch Intern Med*; 2011 Jan 10;171(1):1341–9.
- [9] K.J. O'Leary, D.M. Liebovitz, J. Feinglass, D.T. Liss, D.B. Evans, N. Kulkarni, et al. Creating a better discharge summary: Improvement in quality and timeliness using an electronic discharge summary. *J Hosp Med*. 2009 Apr;4(4):219–25.
- [10] S. Kripalani, A.T. Jackson, J.L. Schnipper, E.A. Coleman. Promoting effective transitions of care at hospital discharge: A review of key issues for hospitalists. *J Hosp Med*. 2007;2(5):314–23.
- [11] A. Mehrotra, J.L. Adams, J.W. Thomas, E.A. McGlynn. The effect of different attribution rules on individual physician cost profiles. *Ann Intern Med*. 2010 May 18;152(10):649–54.
- [12] S. He, G. Gurr, S. Rea, S.N. Thornton. Characterizing the Structure of a Patient's Care Team through Electronic Encounter Data Analysis. *Stud Health Technol Inform*. 2015 Jan;216:21–5.
- [13] The Direct Project | Policy Researchers; Implementers | HealthIT.gov [Internet]. [cited 2016 Dec 9]. Available at: <https://www.healthit.gov/policy-researchers-implementers/direct-project>
- [14] S.N. Thornton, S. He, I. Thraen, M. Hoffman, D.K. Mann, W. Xu. Convening Utah's Healthcare Interests – A Statewide Community Care Coordination Infrastructure. In: AMIA Annual Symposium proceedings. 2016.

Address for correspondence

Shan He, shan.he@imail.org

Using Cloud-Based Physical Activity Data from Google Fit and Apple Healthkit to Expand Recording of Physical Activity Data in a Population Study

André Henriksen^a, Laila Arnesdatter Hopstock^{a,b}, Gunnar Hartvigsen^{c,d}, Sameline Grimsgaard^a

^aDepartment of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway,

^bDepartment of Health and Care Sciences, UiT The Arctic University of Tromsø, Tromsø, Norway

^cDepartment of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway,

^dNorwegian Centre for E-health Research, University Hospital of North Norway (UNN), Tromsø, Norway

Abstract

Large population studies are important sources for medical research. These studies are well planned, well organized, and costly. However, people record health data themselves using different sensors, which are mostly unplanned, unorganized and inexpensive. Nevertheless, self-recorded data might be an important supplement to population studies. The question is how to access and use this data. In the seventh survey of the Tromsø cohort study, questionnaires and accelerometers were used to collect data on physical activity (PA). We now plan to collect historical PA data from these participants, using mobile sensor data already stored in the cloud. We will examine the feasibility of this approach and the quality of this data. Objectively measured historical data will provide valuable insights in the potential and limitations of mobile sensors as new data collection tools in medical research.

Keywords:

Cohort studies; Human Activity; Fitness Trackers

Introduction

The Tromsø Study is a large ongoing population based study in the municipality of Tromsø, Norway. Researchers have conducted repeated surveys 5-8 years apart for a period of 40 years. The Tromsø Study was initiated in 1974 to understand the reasons for the high prevalence of cardiovascular disease in Norway, as well as to find ways to prevent and intervene [1-2]. Since then there have been six additional surveys.

In the most recent survey (2015-2016), the seventh Tromsø Study (Tromsø 7), everyone above the age of 40 was invited. More than 21,000 people participated (65% attendance rate) and went through a wide range of examinations, including questionnaires and interviews, physical and clinical examinations and biological sampling. There are more than 50 different research projects in Tromsø 7, several of which uses PA as a predictor, endpoint or adjustment variable.

In all seven surveys, participants completed a questionnaire, including questions about PA. In Tromsø 7 there were six questions related to PA (sedentary behavior, occupational and leisure-time PA, and PA frequency, duration and intensity).

In the sixth Tromsø Study (2007-08), a subsample of 300 participants wore an accelerometer for seven days, to obtain objectively measured data on PA. Although the coinciding questionnaire was previously validated as a tool to collect data of high-intensity leisure activity, it was also shown that moderate-intensity leisure activity was over-reported when compared to the objectively measured data [3].

In Tromsø 7, 6,300 participants wore the ActiGraph wGT3X-BT accelerometer for one week. A subsample of 700 participants wore the CamNtech Actiwave Cardio for 27 hours. With these sensors, researchers were able to objectively measure PA, energy usage, sedentary behavior, and sleep. The Actiwave Cardio also measured a single lead ECG giving data on heart rhythm and heart rate.

Since PA is an important life style factor, data collected from accelerometers will be important in several ongoing projects. Combining self-reported PA data with objectively measured PA using accelerometers enhances validity in PA measurements. However, having access to continuous PA data over an extended period can be a valuable addition to the data already collected. The next step is to collect historical PA data for participants in Tromsø 7.

Most people carry a smart phone with them throughout the day [4] and a lot of people also have an activity tracker in form of a smart watch or activity bracelet [5]. These devices have several sensors that can measure PA. Android and iPhone users can opt to upload collected health information to Google Fit or Apple HealthKit, which are cloud-based services that allows device vendors and application developers to store health related information in a common online location. This can be beneficial for users, because they can access this information and see details and summaries for all kinds of health related information in one application. A wide range of health related data types can be stored here, but in this project, we will only look into how to access PA and heart rate data.

Compared to the ActiGraph, PA data collected from mobile sensors and connected smart devices are not equally validated for all types of measurements. A systematic review from 2015 indicates that validity of step counting was high, whereas energy expenditure and sleep had a low validity [6]. Validity on PA was inconclusive because of few available studies. In this review, they only considered FitBit and Jawbone, the two leading vendors in the consumer market in 2015. Newer studies seem to indicate the same results [7-8]. Two studies [9-10] from 2016 also indicate that modern wrist worn heart rate monitors are accurate and therefore relevant for our study. Moreover, there will be gaps in the data for participants who do not have a wearable fitness tracker, because not everyone carries their smart phone at all times. These limitations are not necessary drawbacks, but they need to be well understood.

In this vision paper, we will present our plans for retrieving historical PA data, by accessing Google Fit and Apple HealthKit cloud services, for users with smart phones and activity wearables. We will only invite people who participated in Tromsø 7, because their data can enhance existing data sets for Tromsø 7 participants, as well as allow us to examine

the feasibility and quality of this data compared to subjective and objective PA measurements.

Methods

Cloud-Based Health Repositories

On September 17, 2014, Apple released *HealthKit*. Google released *Fit* on October 28, 2014. Users can activate these services to track and store PA, heart rate, diet, sleep patterns and other health metrics. It is possible for users to input information manually, it can be added by third party applications, or it can be automatically collected from internal phone sensors and sensors in connected smart devices (i.e. watches, bracelets, etc.). Users can later view this information or share them with other applications to get a better overview of their health status.

“Google Fit is an open ecosystem that allows developers to upload fitness data to a central repository where users can access their data from different devices and applications in one location” [11]. Google Fit is supported by Android 2.3 and above and consists of four components: 1) the Google Fitness store, 2) the sensor framework, 3) permissions and user controls, and 4) the Google Fit APIs. There are six APIs and the *History API* and the *Sessions API* allows access to historical data and can be used to access relevant information [12-13]. Users can specify what they want to share with the different applications.

Apple HealthKit was introduced in iOS 8 and provides a “centralized, coordinated, and secure data store for health-related information” [14]. There are three ways to access fitness information through the HealthKit API [15]. The *query* approach is appropriate for extracting snapshots of specified periods within the store. These snapshots contain, among other things, aggregated data like step counts.

It is possible to export health data manually from Google Fit, either from the Google web sites or the Google Fit application. Supported formats include TCX-files and CSV-files. TCX-files (Training Centre XML) are a specialized XML format for transferring heart rate, training pace, calories, GPS and other PA related information. Apple HealthKit also supports manual data export to XML-files and CSV-files. Exporting data manually from these services is cumbersome and not a straightforward process for all users. It is however, a way to download the data without writing any code, and we can therefore use it as a quick way to access sample data during development and testing.

We do not know how many Tromsø 7 participants who actually use these cloud services, when they started using them, how often and what type of information they add or how accurate this information is. However, among the more than 21,000 participants in Tromsø 7, it is likely that several have used these cloud services for some time. Furthermore, 83% of the population in Norway have a smart phone [4] and it has been estimated that more than one million smart bracelets and watches will be sold in Norway in 2016 [5]. This is in addition to the 720,000 bracelets sold in 2014 and 2015 [5].

Information Flow

In the proposed solution, we will create a two-part system. In addition, some enhancements on an existing system are required. The first part is a mobile application that will collect and forward health data to the second part. The second part is a Web Service (WS) that will collect, process and forward data to an existing system called EUTRO. EUTRO is the main storage for research data in the Tromsø Study.

Figure 1 shows a simplified illustration of the information flow in this solution. Every time an authorized health application or device receives data, it can also store (1) this data in the Apple HealthKit (A) and/or Google Fit (B) cloud services. This data can be accessed by other authorized applications and the proposed application (C) will be installed on participant's phones, so that this information can be downloaded (2) from the two cloud services. The data will then be uploaded (3) to EUTRO, via the WS, connecting the data to the correct participant in the Tromsø Study.

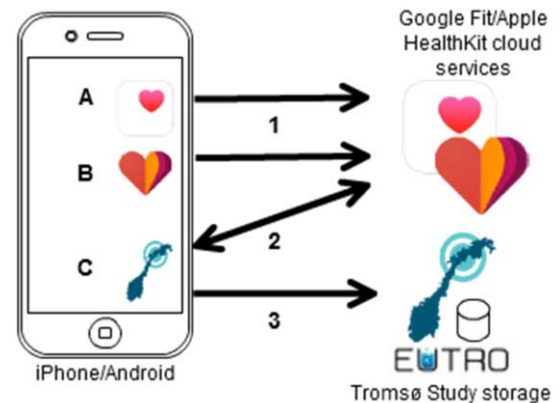


Figure 1-Information flow

System architecture

The proposed system can collect data from potentially thousands of participants, process this data into usable information, and forward it to EUTRO.

Figure 2 illustrates the architecture for this system.

Internal phone sensors and *smart device sensors* generates data which, in addition to manually entered health information, can be shared with *third party health applications* and/or stored in the *Apple HealthKit/Google Fit application* running on the user's phone. The *Apple HealthKit/Google Fit application* will forward this information to the appropriate cloud service. *Third party health applications* can also forward this data to other devices or application specific cloud services.

Participants must install the *Tromsø Study application* on their phones. This application will, upon user authorization, access relevant health data from *Apple HealthKit* and *Google Fit cloud services*. After authorization, the participants must enter a unique number, identifying them as participants in Tromsø 7. This allows linkage of new data to existing data for study participants. Participants will also be able to specify which health data the application can access.

In addition to health data, it is important to have access to sensor and device meta-data. This is necessary in order to examine the data quality, and to examine whether this data collection strategy is a feasible method for collecting objective PA data. After the data transfer to the *Web Service* in the *Tromsø Study backend* is complete, participants can uninstall the *Tromsø Study application*. We are only focusing on historical data in this study and will only download data once. Future PA data is not part of this solution.

The *Web Service* receives the data and processes it into information that can be stored in *EUTRO*. Depending on what data participants have stored in the cloud services, it is possible to store a range of different information. Examples include daily step count, calories burned, heart rate, workouts, sleep patterns

and more. The *Web Service* will forward the processed information to *EUTRO*, where it is stored and made available for active projects in the Tromsø Study.

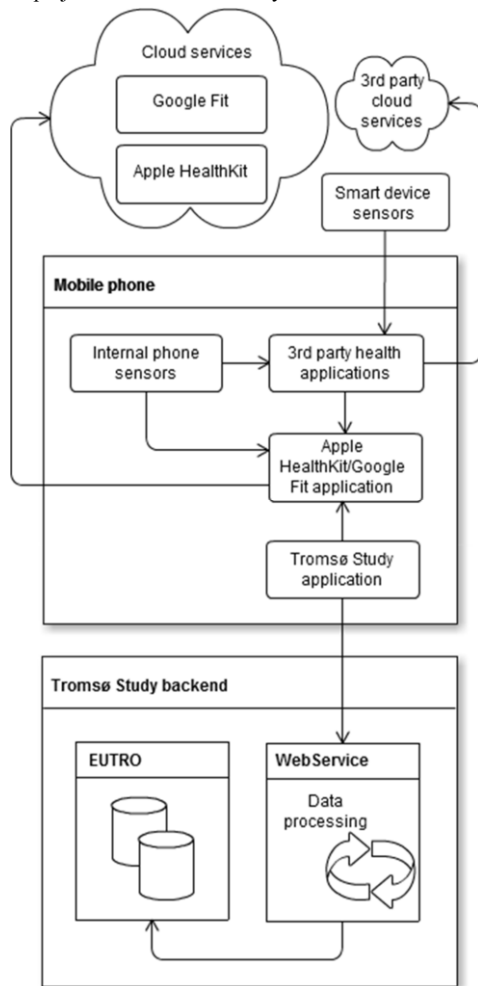


Figure 2-System architecture

We can potentially invite more than 21,000 participants, so it is important that the system scales. Not everyone will connect at the same time, but the amount of available data can be extensive, depending on how much data participants have stored in the cloud and how much they are willing to share.

Mobile application

Our goal is to develop Android and iPhone applications that can read PA and heart rate data from these cloud services. To avoid having to create multiple code bases, we will write the application once and compile it for two platforms. Several solutions allow this. One promising alternative is *React Native* by Facebook [16], which makes it possible to program the application in JavaScript and compile two versions, one for Android and one for iPhone. React Native also supports coding natively, which we will need in order to access the two cloud services.

Through the application, participants can specify which data they agree to transfer, as well as from which period they want to share data. It is not possible to run one query to return everything. Each type of data must be accessed and transferred

individually. It will therefore take some time to collect and transfer the data. Ideally, participants will be willing to share all available data, but because some participants may be reluctant to share everything, they will get the option to specify what they share. Although the amount of data and the level of details for each participant will vary because of this, we hope this will result in a higher participation rate. Figure 3 shows a mock-up of the proposed mobile application, where the user can select what to share and from which period they want to share data.

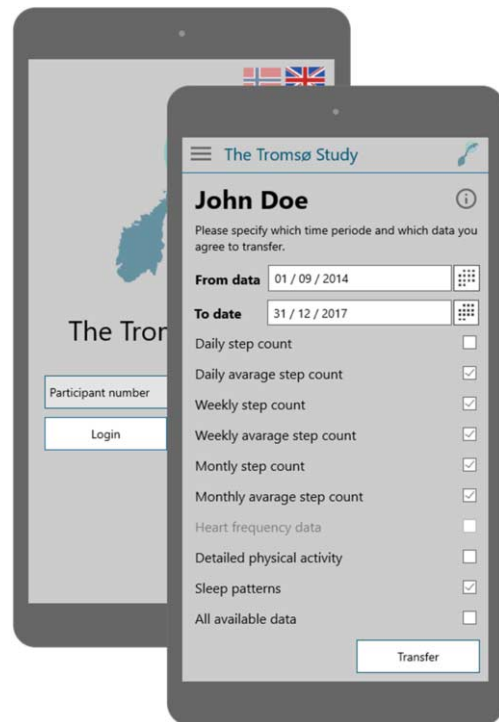


Figure 3-Mobile application mockup

EUTRO

Data collected in the Tromsø Study is stored in EUTRO, a system for managing research projects, biological material, health data and meta data [17-18]. Historical PA data must also be stored here. EUTRO does not currently have an interface for receiving information from mobile applications or other external systems. This means that EUTRO must be extended with an interface to receive data from the new WS, which will serve as an intermediate server for collecting and processing data from the mobile application, as well as a prototype for the changes needed in EUTRO. Until EUTRO is updated, collected data can be manually imported into EUTRO using existing features.

Web Service

The mobile application's only function is to transfer health data from the two mentioned cloud services. The planned WS will be the most complex part of the system. This is because the transferred data will be of different types, have different detail levels and be from different periods. This data must be processed into usable and comparable information.

The data that is most likely to be available is step count data, which might be aggregated hourly, daily, weekly or monthly, or averaged over a longer period.

The following types of health data will be collected, processed and forwarded, if they are available:

- Daily, weekly and monthly step counts
- Daily, weekly and monthly average step counts
- Daily, weekly and monthly average heart rate
- Sedentary behavior
- Sleep patterns
- Exercise bouts, including detailed heart rate, intensity, average speed and inclination
- Estimated precision on each data set
- Sensor meta data

Security is important when working with health data. Information transferred between systems will be encrypted using SSL. Authentication is achieved by using a unique identification number included in the invitation letter. When the mobile application has completed the data transfer, it will call a method to disable the identification number. This will further enhance security. The WS will be REST-full, that is, stateless and self-contained, encouraging a simple, lightweight and fast interaction between server and client. This is important to support the high number of potential participants.

Challenges

There are several challenges in this project. Motivating participants to install the application in order to share their information, is probably going to be the main challenge.

Wearing an ActiGraph, a standalone physical device, not connected to anything is one thing. With the ActiGraph, it was expected to be challenging to motivate participants to wear the device for several days, because of the inconvenience it might cause. However, compliance among participants in Tromsø 7 was high, with a 94% acceptance rate. Installing an application with the purpose of sharing personal health related data is another thing altogether. There are several challenges with this approach where trust and motivation probably are the two most important.

Installing any application on your personal phone, where you have so much personal data stored, requires a certain level of trust. There are many fraudulent applications which only function is to hijack your phone and access your personal information. As a participant, you must trust that the application only does what it claims it will do. Specifically, that the application will only access data the participant has agreed to share, and that the system is secure and data interception is impossible during transfer.

We need to find ways to motivate participant to share their data. The Tromsø 7 survey contained questions, tests and measurements, divided into two phases. Those attending both phases have already spent about 4 hours contributing data to Tromsø 7. Participants in the PA project, wearing the ActiGraph for a week, have contributed even more. Asking participants to use more time can potentially be a challenge. However, because the effort needed to share data is very low, once the application is installed, we are confident that many will be willing to participate in this extension to the PA project. After installing the application, participants only have to input an identification number and select which data to share. The rest is automatic and requires no additional input.

However, installing the application and using it properly can be a challenge for some participants. In a future survey, it is possible to set up a station where people can get assistance with the setup. Because the data collection phase of Tromsø 7 is completed, there is no infrastructure to provide such assistance now. This will be resolved by making a detailed instruc-

tion that must be included in the invitation letter. In addition, making an instruction video might result in some additional participants.

Ethics

Actively installing the application and selecting which data to share requires informed consent from all study participants. We will apply for recommendation from The Regional Committee for Medical and Health Research Ethics, and for approval from the Norwegian Centre for Research Data on the processing of personal data.

Expected Results

Scenarios

The most likely scenario is a person who only has a mobile phone and no wearable smart device. For this person, only a limited set of data will be available, and only in periods the person has the phone with him/her. For several hours of the day, as well as during the night, the internal mobile sensors will not collect any activity data. In this scenario, we expect that the system will get access to daily step count, which is unlikely to be accurate because the phone is not carried around all day.

A more data rich scenario will be a person that has an activity bracelet or watch that also can collect heart rate. For this person, it will be possible to collect heart rate data for specific days, as well as to compare this data with different levels of PA. If this person brings a GPS-enabled phone during workouts, or has GPS built into the bracelet/watch, it will be possible to access average speed, max speed, altitude differences, average heart rate, peak heart rate and more. As an example, measuring how the heart rate recovers after exercising is an indicator of cardiovascular health.

Data Usage

The information collected from participants in Tromsø 7 is extensive. All study participants completed comprehensive questionnaires, sharing information about their lifestyle, general health, medical history, drug usage, diet, alcohol consumption, smoking, education, social status, PA and other areas. Measurements of height, weight, hip-waist, blood pressure, heart rate, SpO₂ and pain sensitivity were also collected. Biological samples of blood, saliva, urine, feces and samples from the nose/throat were also taken. A subsample underwent dental examinations, ECG-recording, cognitive tests, physical function tests, carotid ultrasound, eye tests, lung function tests, measurements of bone densitometry and body composition (by DEXA), echocardiography, and heart-, lung- and carotid auscultation. Many of the participants also attended earlier surveys in the Tromsø Study, which gives the possibility to analyze repeated measurements, enriching the data further.

Because the validity of the data collected by wearable devices is unknown or low for most data types related to PA, step count data will be collected first. Heart rate data for participants using heart rate monitors will also be collected since these sensors are considered accurate. Participants who already wore accelerometers in Tromsø 7 will be invited first. With these participants, existing accelerometer data can be compared with participant's phone and wearable sensors collected in the same period. This will help determine how accurate step count and heart rate data collected from these sensors are. If accurate, all remaining participants in Tromsø 7 will be invited. This will potentially result in objective measurements for step counts for a proportion of the 21,000 participants in Tromsø 7.

Discussion

PA is an important lifestyle factor and is relevant in multiple projects in Tromsø 7, both as a predictor, endpoint and adjustment variable [19]. Data from accelerometers include heart rate, PA, sleep patterns, energy usage and sedentary behavior. However, only one week of PA data was collected using accelerometers. To get a broader picture of PA in a population, it is of interest to include historical PA data stored in Google and Apple cloud services. Data from accelerometers used in Tromsø 7 is considered accurate. This level of accuracy is, for the most part, not achievable through mobile and smart device sensors.

Participants will have a wide range of mobile devices, resulting in data from a wide range of different sensors. For those wearing a smart watch or bracelet with activity and pulse sensors, there will also be a lot of variety in how accurate these sensors are. Until data collection starts, the accuracy of the data is unknown. In addition, from participants that do not have a smart watch or activity bracelet, activity data will only be available for parts of the day. All these limitations are okay, if we know about them. The data should not be used alone, but rather as supplements to the data already collected in Tromsø 7.

Population based studies are valuable sources for new knowledge and for monitoring the status and development in population health and for disease risk factors. However, these types of studies are costly, and epidemiologists are concerned that participation rates are declining both in Norway and internationally. Hence, we need to motivate study participation and develop new tools for data collection. Smart phones come forward as promising data collection tools. They are prevalent, it is easy for study participants to donate their data, and large amounts of data can be collected from many study participants at a low cost.

Physical inactivity is an emerging and important disease risk factor in western populations. We aim to develop solutions to access data on PA from smartphones using internal and connected sensors, and to examine the strengths, possibilities and the potential of smart phones as data collection tools for future medical research on PA.

Future Work

Collecting historical data is only the first step in including PA and heart rate data from mobile sensors in the Tromsø Study. The next step will be to collect the same kind of data for future activity. That is, to collect all PA and heart rate data for several months using mobile sensors and sensors in smart watches, bracelets and other smart devices. The experience from accessing historical data will be valuable going forward. After implementing the system described in this paper, we will expand the system design into a solution for collecting future data.

Conclusion

In this paper, we described our plans for collecting up to two years of historical PA and heart rate data from participants in the Tromsø 7 survey. Accessing this data is possible through Google Fit and Apple HealthKit cloud services. Access is restricted, and participants must consent to donate their data to the Tromsø Study database. The quality of the data is unknown and we therefore wish to examine it. The data will be a supplement to existing PA data from accelerometers and questionnaires in Tromsø 7. In order to validate these sensors as potential new tools for objectively measure PA, we will com-

pare data collected from mobile sensors with existing accelerometer data.

References

- [1] Jacobsen, A.E. Eggen, E.B. Mathiesen, T. Wilsaard, and I. Njølstad, Cohort profile: The Tromsø Study, *International Journal of Epidemiology* **41** (2012), 961-967.
- [2] I. Njølstad, E.B. Mathiesen, H. Schirmer, and D.S. Thelle, The Tromsø study 1974–2016: 40 years of cardiovascular research, *Scandinavian Cardiovascular Journal* (2016), 1-6.
- [3] A. Emaus, J. Degerstrom, T. Wilsaard, B.H. Hansen, C.M. Dieli-Conwright, A.-S. Furberg, S.A. Pettersen, L.B. Andersen, A.E. Eggen, L. Bernstein, and I. Thune, Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromsø study, *Scandinavian Journal of Public Health* **38** (2010), 105-118.
- [4] Statista, Percentage of people who use a smartphone in Norway from 2012 to 2016 in, 2016.
- [5] Tek.no, Hele 800 000 nordmenn forventes å kjøpe et smartbånd i år, in: Tek.no, 2016.
- [6] K.R. Evenson, M.M. Goto, and R.D. Furberg, Systematic review of the validity and reliability of consumer-wearable activity trackers, *The International Journal of Behavioral Nutrition and Physical Activity* **12** (2015), 159.
- [7] R.E.R. Reid, J.A. Insogna, T.E. Carver, A.M. Comptour, N.A. Bewski, C. Sciortino, and R.E. Andersen, Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment, *Journal of Science and Medicine in Sport* (2016).
- [8] E.E. Dooley, N.M. Golaszewski, and J.B. Bartholomew, Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices, *JMIR Mhealth Uhealth* **5** (2017), e34.
- [9] S.E. Stahl, H.S. An, D.M. Dinkel, J.M. Noble, and J.M. Lee, How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough?, *BMJ Open Sport Exerc Med* **2** (2016), e000106.
- [10] M.P. Wallen, S.R. Gomersall, S.E. Keating, U. Wisloff, and J.S. Coombes, Accuracy of Heart Rate Watches: Implications for Weight Management, *PLoS One* **11** (2016), e0154420.
- [11] Google Inc., Fit SDK overview, in, 2016.
- [12] Google Inc., Fit APIs, in, 2016.
- [13] Google Inc., Fit History API, in, 2016.
- [14] Apple Inc., Introduction to HealthKit, in, 2016.
- [15] Apple Inc., HealthKit API Reference, in, 2016.
- [16] Facebook Inc., React Native, in, 2016.
- [17] UiT - Department of Community Medicine, Eutro, in, 2015.
- [18] UiT - Department of Community Medicine, Eutro, in, 2011.
- [19] UiT - Department of Community Medicine, Hva forsker vi på?, in, 2016.

Address for correspondence

André Henriksen, Department of Community Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

E-mail: andre.henriksen@uit.no

Making Sense of Patient-Generated Health Data for Interpretable Patient-Centered Care: The Transition from “More” to “Better”

Pei-Yun Sabrina Hsueh^a, Sanjoy Dey^a, Subhro Das^a, Thomas Wetter^{bc}

^a Center for Computational Health, Watson Research Center, Yorktown Heights, New York, USA

^b Department of Biomedical Informatics and Medical Education, University of Washington, USA

^c Institute of medical Biometry and Informatics, University of Heidelberg, Germany

Abstract

The rise of health consumers and the accumulation of patient-generated health data (PGHD) have brought the patient to the centerstage of precision health and behavioral science. In this positional paper we outline an interpretability-aware framework of PGHD, an important but often overlooked dimension in health services. The aim is two-fold: First, it helps generate practice-based evidence for population health management; second, it improves individual care with adaptive interventions. However, how do we check if the evidence generated from PGHD is reliable? Are the evidence directly deployable in realworld applications? How to adapt behavioral interventions for each individual patient at the touchpoint given individual patients' needs? These questions commonly require better interpretability of PGHD-derived patient insights. Yet the definitions of interpretability are often underspecified. In the position paper, we outline an interpretability-aware framework to handle model properties and techniques that affect interpretability in the patient-centered care process. Throughout the positional paper, we contend that making sense of PGHD systematically in such an interpretability-aware framework is preferable, because it improves on the trustworthiness of PGHD-derived insights and the consequent applications such as person-centered comparative effectiveness in patient-centered care.

Keywords:

Informatics; Patient-Centered Care; Machine Learning

Introduction

Owing to the trends in participation health [1] and value-based care [2], health consumer perception and adoption of direct-to-consumer devices and sensors and citizen science are at all-time high.[3] Meanwhile, more and more leading healthcare systems are evaluating the secondary use of patient-reported outcomes measures (PROM) in electronic health records (EHR).[4] A plethora of patient-centered data generating devices and care processes are producing masses of data.

The early evidence has started to emerge and stimulate the field through best practices.[5] As shown in Figure 1, many of the healthcare applications hinge on the convergence of PGHD and clinical data, as well as the clinical and patient information system. This is in line with the vision of patient-centered care as defined in a recent patient advocacy testimonial in *Health Affairs* [6] as “the experience (to the extent the informed, individual patient desires it) of transparency, individualization, recognition, respect, dignity,

and choice in all matters, without exception, related to one's person, circumstances, and relationships in health care.”

In particular, two opportunity areas emerge. First, PGHD helps generate practice-based evidence for population health management.[7] Traditionally, clinical evidence is generated from costly randomized controlled trials (RCT) and comparative effectiveness studies in observational clinical data such as EHR, claims and administrative databases. However, as indicated in the recent Institute of Medicine (IOM) report [8], it is important to start designing clinical information systems that can help capture the patient's state such as social, behavioral and environmental determinants, while fitting situational use of PGHD-derived evidence in the clinical context.[4,6]

Second, PGHD improves individual care through adapting interventions against the incoming stream of patient observations (e.g., lifestyle and physiological measures) and outcome history. The initial results show potential in making sense of PGHD for clinicians and care coordinators [9,10] and fitting situational use of PGHD (for example, for adaptive trials of mobile app-based behavioral interventions [11]).

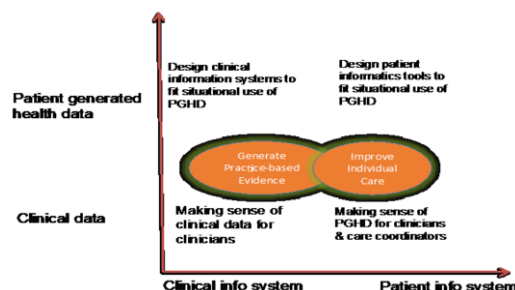


Figure 1– Convergence between PGHD and clinical data in clinical/patient information system for patient-centered care.

Both opportunity areas lead to many subsequent questions regarding the interpretability of PGHD in real-world healthcare applications. For example, how do we check if the chunks of evidence generated from PGHD are reliable? Are they directly deployable in real-world healthcare applications? How to adapt the behavioral interventions for each individual patient at the touchpoint given individual patients' needs? Despite the recent attention in developing interpretable machine learning models for healthcare applications [12,13], the definitions of interpretability are underspecified due to the many different motivations.[14]

Method

In this positional paper, we hereby first review the major dimensions underneath the interpretability mentions. Then, we address the next challenge for the development of health informatics tools to enhance the interpretability-awareness of PGHD-derived insights. We summarize the various issues and model properties that should be addressed in an interpretability-aware framework (as shown Figure 2 below). Next, we highlight the emerging practices in which the framework provides value to patients and clinicians and improves care delivery. Finally, we examine present and future challenges to incorporating PGHD-based evidence back into the care flows, and utilizing cleansed and approved data for purposes beyond their primary context and motivation of collection.

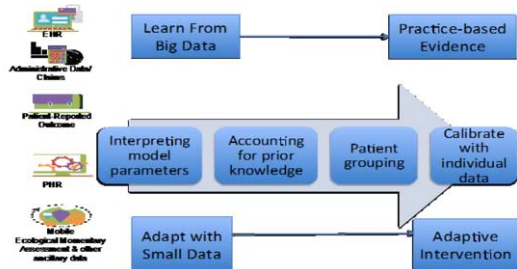


Figure 2—Interpretability-aware framework for population evidence discovery and individual intervention adaptation

Summarizing Major Dimensions of Interpretability from Patients' Perspective

To understand major dimensions behind interpretability, we need to first review the motivations of improving interpretability in different dimensions. For example, [15] explains and quantifies the interpretability metrics for decision rule-based analytics. Here are a few dimensions that we identified for applying machine learning models to generate practice-based evidence and adapt interventions for individuals.

Complexity: As noted in [16], humans are best able to reason about models that are composed of simple forms such as decision trees.[17,18] This interpretability constraint on model complexity persists when it comes to generating N-of-1 models to tailor intervention recommendations. In practice, this involves further constraining policy learner architectures to limited sets of simple rules – first at the population level, and then at the individual level. This has inspired a variety of methodological advancements recently. For example, combining reinforcement learning and regression trees can produce simpler policies for human inspection.[19] Recent results suggest that simple explainable policies could be achieved in complex problems, such as fine-tuning adaptive evaluation of behavioral intervention strategies.[11]

Knowledge Structural Similarity: The interpretability can also be defined as a distance metric based on the known domain knowledge. Intuitively, closer the concepts covered by PGHD, easier is their interpretation. Suppose there exist some relationships among the features defined in a hierarchical manner as represented by a tree or directed acyclic graph. The inherent hypothesis is to make sure that the similar features should have similar model co-efficients learnt by the model. One popular approach is group LASSO [20] based techniques,

which can pose additional constraints on the above equation to make sure that all the nodes rooted in a particular subtree will have similar parameters during model learning.

Quality: Another major hindrance of patient interpretation of data is the quality issues of PGHD. The sources of errors in PGHD is multi-fold. For self-reported data, the baseline of patient-reported outcome and lifestyle information (for example, daily calorie intake) varies from person to person, resulting in under-reporting or over-reporting. Moreover, the measurements from health wearables and IoT (Internet-of-Things) sensors are often noise corrupted due to inappropriate placement, incorrect use, regular wear and tear of the devices. To overcome the data quality issues, the process of imputing missing values and detecting anomalies and outliers are then needed to train models; which in turn is used for forecasting.

Usability: HCI researchers conducted qualitative studies (e.g., [21]) to make sense for patients of their own data and in addition, to identify interpretability-impeding factors, including: *confounders*, *noisy on meaningful and irrelevant measures*, and *how to determine the time lag of outcome-affecting triggers at the individual level*.

Causality: To further ensure that the framework can provide statistically sound interpretation, we surveyed the causal inference research to understand the effect of following assumptions [22]: *consistency* (i.e., whether one's features and outcomes are consistently observed when the actions are taken), *stability* (i.e., whether one's features and outcomes are affected by other subjects' actions), and *unmeasured confounders* (i.e., what is the sensitivity of conclusions with factors that influence the assignment of treatments).

The survey leads us to believe that there exist quite diverse views behind the concept of interpretability, and the field of health informatics needs to reconcile the differences by first making the previously implicit assumptions more explicit.

Developing Interpretability-Aware Framework

The development of the interpretability-aware framework is based on a two-layer approach: (1) “Learn from Big data”: PGHD from heterogeneous sources are aggregated to learn practice-based evidence for optimal outcome (e.g., efficiency of patient capacity) and (2) “Adapt with Small data”: interpreting evidence through comparing the effectiveness across interventions and adapting in a patient-centered way.

Learn from Big Data: PGHD to Practice-based Evidence

First, in order to learn interpretable practice-based evidence that can be conferred from the secondary use of PGHD, we need principled and scalable approaches to address the interpretability issues: model complexity and knowledge structural similarity. The goal is to minimize the model complexity during the process of interpreting model parameters, while boosting knowledge structural similarity to account for prior knowledge during model development.

Interpreting Model Parameters: Most informatics tools rely on learning a set of parameters that are associated with the features extracted from raw PGHD. In particular, an objective loss function is defined based on the original model outcomes and the predicted outcomes using the parameters of the model and then, those parameters are learnt from PGHD such that the loss function is minimized. Let X denote the original feature set, y the outcome, and W the model parameters. The loss function can be represented as below:

$$\min_w L(X, W, y)$$

Once the model parameters W have been learnt from the model, it is further analyzed to interpret the model. For example, the coefficients obtained from a logistic regression model can be converted into an odds ratio, which is easier to interpret by domain experts who prefer models of simpler forms.[16]

Another useful technique to interpret the model parameters is to impose some sparsity constraints on the model parameters. In that case, the objective function will contain both the original loss function and an additional penalty imposed on the complexity of the parameters. The function is shown in the equation below, where $\lambda(W)$ denotes the complexity of model parameter set W . One popular example of such loss function is L1-norm regularization penalty [23], since it can perform feature selection simultaneously with model learning, and that will help reduce model complexity:

$$\min_w L(X, W, y) + \lambda(W)$$

Taking prior knowledge into account: Despite that PGHD are usually collected from heterogeneous data sources, the observational data can be interrelated by certain latent factors or well-established medical knowledge. For example, most of the interventions and care workflows are conducted using a few well-established guidelines. Such guidelines can capture the inherent relationships among observational healthcare data including both PGHD and EHR.

In addition, features obtained from observational data may be well structured with semantic relationships among them. For example, drugs, adverse reactions and diagnostic similarity each has clear role in a hierarchical organization based on how specific or generic it is in its mechanism of actions.

These pre-existing relationships are usually curated with the help of multiple domain experts and a standardized protocol.[24] In particular, the interpretability metrics is quantified as knowledge structural similarity, using a distance metric based on the distance of each pair of features in its hierarchy

$$H_{ij} = \frac{\text{depth}(\text{LCA}(X_i, X_j))}{\max(\text{depth}(X_i), \text{depth}(X_j))}$$

Here, LCA defines lowest common ancestor of each pair of two features X_i and X_j , and depth of X_i defines length of the shortest path from the root of the tree toward X_i . Moreover, this distance metric is normalized by the maximum depth of the two features. Finally, this new metric is incorporated in the original penalty structure of the objective function as below:

$$\min_w L(X, W, y) + \lambda_1(W) - \lambda_2(H)$$

Note that similar concepts can be generalized when prior relationships exist not only among the features from one type of data, but also features coming from multiple data-sources such as the heterogeneous exogenous determinants in PGHD.

Adapt with Small Data: Interpret from Patient Perspective

Second, in order to adapt interventions on an individual basis, we include an optimal policy learning component that can tailor interventions against incoming streams of “small” data. The goal is to provide interpretable evidence that can help patients and their care teams make decisions that meet patients’ individual needs. In this paper, we introduce components of patient grouping and calibration to evaluate and to inform interpretability-aware analytics.

Patient Grouping: To overcome the barrier of “one-size-fits-all” guidelines to treat all patients as an “average” patient, we apply behavior segmentation methods [25] that can identify

sub-cohorts that exhibit distinctive behavioral differences and extract signature behavioral patterns. In the framework, behavioral factors are constructed as a composite of multi-source features for each subject in the cohorts. We illustrate the design of a framework able to generate, analyze and re-rank the risk factors for the behaviorally different segments (as shown in Figure 3; for a more detailed description of the key component of patient grouping, please refer to [25]).

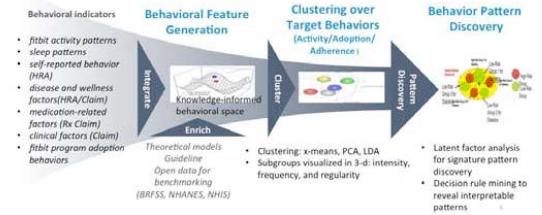


Figure 3 – Patient grouping for subcohort identification and behavioral pattern discovery

While patient grouping finds common static subcohorts directly against the incoming patient data, it has its own disadvantage. For example, it ignores the temporal relation within and among patient behaviors. The behavioral pattern discovery approach applies dynamic item response analysis to incorporate extensions to temporal patterns across multiple streams of data.[26] Such techniques are able to find patterns within groups of variables, patterns in a sequence of variables, or both. This further confirms the need to not treat all patients as an “average” patient – as defined in the guidelines.

In order to learn how confident we can be to deploy the PGHD-derived evidence in certain sub-cohorts, we need to quantify the population representativeness in each sub-cohort. This is achieved through computing an entropy-based similarity index between the sub-cohorts. Low similarity shows the separability of the population segments.

To improve the interpretability, prototypical user examples are used to provide explanations of why a certain evidence is generated for a particular sub-cohort. For each sub-cohort, we compute the prototypical core among the patients, as defined as 10% of the patients closest to the centroid in each patient group. The population means of the core and the whole sub-cohort are then compared with F-statistics. Further the features whose means are beyond two standard-deviations from the entire population means, constitute the distinctive features of that segment. These then become parts of the explanation.

Calibration against Individual Data: In addition, we further design an adaptive real-time approach to address the quality issues abovementioned and to provide intermittent feedback to individual patients.

First, in the data cleansing step, the PGHD is processed for missing values and outliers using the machine learning methods and knowledge of domain experts.

Next, the data goes through an adaptive predictive model, where the underlying data-driven phenomena are modeled as a noisy dynamical system. For each individual patient, predictions of future outcomes are compared with outcomes that are actually measured. Low covariance variables are then marked for correction. A recalibrated model is then built from the original model and the correction feedback. The correction metric for the erroneous variable is then fed back to the data processing step to calibrate the individualized systems.

Further, the system can prompt the users for labeling of inputs corresponding to the erroneous variables. This kind of dynamic feedback system will improve the data quality for analysis and improve the prediction accuracy for each person.

Discussion

In this positional paper, we summarize the major dimensions of interpretability and describe the interpretability-aware framework for further adding a patient focus into the care process. The interpretability-aware framework helps foster a continuous learning health care system as pictured in the “All of US” platform under the Precision Medicine Initiative (PMI).[27] The framework will establish prospective effectiveness based on the basic phenotypes found in mass data and new cases matching some basic phenotypes. This framework, when coupled with the best practice defined for clinical flow to increase patient understanding, is expected to further fuel the patient-centered care model for minimally disruptive medicine.[28] The framework is expected to facilitate the integration between science of data and science of care at the touchpoint. This is especially important for the complex care scenarios wherein standard guidelines and general population-based evidence fall short.

Comparing person-centered effectiveness at touchpoint

Traditionally, comparative effectiveness studies are used to provide evidence to handle “average” patients. Comparative effectiveness studies, if done with a patient focus, can empower patients to better understand and take charge of their decisions. Therefore, in the proposed framework, we further include additional patient grouping and re-calibration steps to reinforce the patient focus.

In retrospect, it can also help identify hypotheses to be verified or falsified. Although classification modeling analysis has become a routine tool in health informatics research, extracting actionable insight from such information remains a major challenge. Formalizing the interpretability metrics and framework such as the correction metrics can help pinpoint the previously unobserved inefficiency of practice and attribute it to variables that matter to patients sub-cohorts and providers, as opposed to those that only add noise.

Putting Big Data and Small Data Together

In terms of the actual implementation, formalizing and scaling up the interpretability-aware framework means solving various practical problems. These include conducting feasibility study of monitoring devices and developing new forms of outcomes.

Take the assessment of stress-behavior relationship as an example. Traditionally, this is done with survey-based ecological momentary assessment (EMA).[29] By coupling EMA with mobile devices, we can repeatedly collect exposure data of psychosocial stressors in ecologically valid settings such as home and work, and in real time. Compared to the survey-based EMA, mobile EMA enables collecting data with the immediate context and substantially reduces recall bias.[30]

In fact, oftentimes, we would not need to develop individual models from scratch, but rather to recalibrate the existing population-based model using user’s own data. The confidence level can be estimated with the sub-cohort identified from the patient grouping step. The recent trends of N-of-1 trials have started to provide evidence on the effectiveness of such approaches for adaptive design.[30,31] Currently in the field, researchers are attempting to apply N-of-1 methods to

develop individualized predictive pathways that can be applied to adapt interventions at the touchpoint directly.

The societal approval and ethical issues going forward

It will also require a cultural shift from large, population-based trials to ad-hoc, post-trial analyses that aim to interpret the factors that cause some patients to be responsive. We have to meet such challenges as balancing the economic power of holding the data against the moral maxim of equitable access for citizens to individualized recommendations. Since the data themselves often are public domain there is a moral obligation to make them available free of charge. Since the aggregation of the data requires proprietary technology, there is equal legitimacy to charge market prices for such recommendations.

This compares to the question whether medical advice is a commodity or a merchandise. We have to find answers for patients who find themselves in a Gestalt whose treatment by far exceeds their individual economic reach. We have to find agreement, world wide—since the origin of data is world wide—how to handle the detection in passing of alarming conditions: are we entitled to know the name of data source and its identity and to take action? May we, who detect a risk, even be obliged to take action or else be taken accountable for nonfeasance, as physicians would be? Therefore, do we, as informaticians who are engineers in a wide sense, assume the role of physicians and hence inherit the moral standards that physicians have to uphold? Does such an obligation differ between individual risks and societal risks such as epidemics?

Today, we are far from answers. We will, however, demonstrate how principles and maxims from ethics can be used as tools to address these and more question.[32]

Conclusion

The advent of large-scale PGHD collected from diverse sources poses unique opportunities to harvest insights about patients’ behavior and response to a particular treatment, which can ultimately be used to derive knowledge for making better clinical and self-care decisions. Given the prevalent adoption of EHR and the shift to value-based care, many leading healthcare systems are now evaluating the value of PGHD generated from care processes, which can be either used directly to enhance care processes, or aggregated in big data to derive practice-based evidence.

Traditionally, the generation of clinical evidence relies on RCT and comparative effectiveness studies, which employ costly clinical trial research design and observational data sources such as EHR, claims and administrative data. While these data sources reveal valuable information about treatment effect and healthcare utilization, individual patient data are still needed to identify individually outcome-differential health determinants such as in their social, behavioral, environmental, and psychological factors.

However, the current interpretability of PGHD-driven insights is still questionable when applied to real world applications. In addition, the interpretation of such insights often hinges on experienced healthcare professionals (e.g., care managers). By applying the framework to communicate, it is easier to advocate for post-statistics decision models, which often incur new forms of outcomes and social desirability bias.

This inevitably incorporates humans in the loop during the process of transforming big data to practice-based knowledge and deploying such evidence at the touchpoint for individuals.

In this positional paper, we first identify the major dimensions of interpretability. Then, we depict our interpretability-aware framework in which interpretable analytics are enhanced to better incorporate PGHD insights back to the care flow. The aim is two-fold.

First, the proposed framework gives care team tools to address two major challenges: (1) Enable the generation of practice-based evidence from aggregating “big data” from heterogeneous PGHD sources, especially useful for complex care scenarios wherein no clear evidence or guidelines are applicable; (2) Learn to further adapt population health-based recommendation with “small data” from individuals or sub-cohorts for self-care and experimentation.

Second, the framework aims to make personalized recommendations on *what to intervene on* at the touchpoint. However, oftentimes the human opinions are subjective and highly depend on their prior expertise and training, which lead to underlying bias and noise factors and decrease the generalizability of the model. Therefore, how to extract best practice and make it scalable throughout organizations and in real world applications is the key for future uptake in practice.

These issues can be mitigated to some extent by taking human knowledge into account in an earlier stage of model development rather than during the model validation step. Therefore, the proposed interpretability-aware framework can help pinpoint more relevant yet explainable risk factors along with their relationships with disease outcome. The bottom line is that these methodological advances should be used to augment, but not to replace the central role of human insights in predicting behavior that can be intervened on.

Using the interpretability aware framework, we can also establish prospective effectiveness of interventions based on more interpretable basic phenotypes found by the patient similarity in mass data. This is important to identify evidence deployable for individuals with uncommon characteristics rather than hiding the apparent noise they contribute as variance and noise.

References

- [1] K. Gray, F.J. Martin-Sanchez, G.H. Lopez-Campos, M. Almalki, and M. Merolli, Person-generated Data in Self-quantification. *Methods of Information in Medicine* 55:6 (2016).
- [2] Catalyst for Payment Reform, National Scorecard on Payment Reform. CPR, Berkeley, California, 2014. Available from: <http://www.catalyzepaymentreform.org/how-we-catalyze/national-scorecard/89-how-we-catalyze/scorecard-and-compendium/126-scorecard-on-medicare-payment-reform> [last visit April 22 2017]
- [3] J.A. Guest, L. Quincy, Consumers Gaining Ground in Health Care. *JAMA* 310:18 (2013), 1939–40.
- [4] A.W. Wu, H. Kharrazi, L.E. Boulware, and C.F. Snyder, Measure once, cut twice—adding patient-reported outcome measures to the electronic health record for comparative effectiveness research, *Journal of Clinical Epidemiology* 66:8, (2013), S12–S20.
- [5] P.S. Hsueh, Y.-K. Cheung, S. Dey, K. Kim, F. Martinez-Sanchez, S. Petersen, and T. Wetter, Added Value from Secondary Use of Person Generated Health Data in Consumer Health Informatics., *IMIA Yearbook*, 2017.
- [6] D.M. Berwick, What “patient-centered” should mean: confessions of an extremist, *Health Affairs* 28:42009;: w560.
- [7] T. Coelho, A patient advocate’s perspective on patient-centered comparative effectiveness research, *Health Affairs* 29:10 (2010), 1885–90.
- [8] Institute of Medicine, *Capturing Social and Behavioral Domains and Measures in Electronic Health Records*. Washington, D.C.: National Academies Press, 2014.
- [9] S.S. Gollamudi, E.J. Topol, and N.E. Wineinger, A framework for smartphone-enabled, patient-generated health data analysis, *PeerJ* 4 (2016), e2284.
- [10] L. Aschettino, K. Baldwin, B. Friedman, R. Grady, L. Grebner, M.E. Hennings, and L. Washington, Including Patient-Generated Health Data in Electronic Health Records, *Journal of AHIMA* 86:2 (2015), 54–7.
- [11] X. Hu, P.S. Hsueh, C.Chen, M. Qian, F.E. Parsons, I. Ensari, K. Daz, and Y.K. Cheung, An interpretable health behavioral policy for mobile device users, *IBM Journal of Research and Development*, 2017.
- [12] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, Intelligent models for healthcare: Predicting pneumonia risk and hospital 30-day readmission, *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*, 2015
- [13] K.R. Varshney and H. Alemzadeh, On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products, 2016. Retrieved from <http://arxiv.org/abs/1610.01256>.
- [14] Z.C. Lipton, The Mythos of Model Interpretability, *Proceedings of ICMML Workshop on Human Interpretability in Machine Learning (WHI 2016)*, 2016,
- [15] H. Lakkaraju, S.H. Bach, J. Leskovec, Interpretable Decision Sets. *KDD '16* (2016), 1675–1684.
- [16] J. Feldman, Minimization of boolean complexity in human concept learning, *Nature* 407:6804 (2000), 630–633.
- [17] M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Statistical Methodology* 68:1 (2010), 49–67.
- [18] S. Dey, K. Lim, G. Atluri, A. MacDonald, M. Steinbach, V. Kumar, A pattern mining based integrative framework for biomarker discovery, *Proc. of the ACM Conf. on Bioinformatics, Computational Biology and Biomedicine*, 2012.
- [19] M. Sridharan and G. Tesauro, Multi-agent q-learning and regression trees for automated pricing decisions, *Game Theory and Decision Theory in Agent-Based Systems* (2002), 217–34.
- [20] R. Tibshirani, Regression shrinkage and selection via the lasso: a retrospective, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73:3 (2011), 273–282.
- [21] R. Karkar, J. Zia, R. Vilardaga, S.R. Mishra, J. Fogarty, S.A. Munson, and J.A. Kientz, A framework for self-experimentation in personalized health, *JAMIA* 23:3 (2016), 440–448.
- [22] J.M. Robins, A. Rotnitzky, and L.P. Zhao, Estimation of regression coefficients when some regressors are not always observed, *Journal of American statistical Association* 89:427 (1994), 846–66.
- [23] B.L. Westra, S. Dey, G Fang, M Steinbach, V Kumar, C Oancea, et al., Interpretable predictive models for knowledge discovery from home-care electronic health records, *Journal of Healthcare Engineering* 2:1 (2011), 55–74.
- [24] S. Dey, G. Simon, B. Westra, M. Steinbach, and V. Kumar, Mining Interpretable and Predictive Diagnosis Codes from Multi-source Electronic Health Records, *Proceedings of the 2014 SIAM International Conference on Data Mining*, 2014.
- [25] P.S. Hsueh, H. Chang, and S. Ramakrishnan, Next-Generation Wellness: A Technology Model for Personalizing Healthcare. In: Weaver CA, Ball M, Kim G, Kiel JM (eds.), *Healthcare information Management*, 4th Edition, Springer Publishing, 2016.
- [26] P.S. Hsueh, X. Zhu, V. Deng, S. Ramarishnan, and M. Ball, Dynamic and accretive composition of patient engagement instruments for personalized plan generation, *Studies in Health Technology and Informatics*, 2014.
- [27] F.S. Collins and H. Varmus, A new initiative on precision medicine, *The New England Journal of Medicine* 372:9 (2015), 793–5.
- [28] Leppin, A., Montori, V., & Gionfriddo, M. (2015). Minimally Disruptive Medicine: A Pragmatically Comprehensive Model for Delivering Care to Patients with Multiple Chronic Conditions. *Healthcare* 3:1 (2015), 50–63.
- [29] S. Shiffman, A.A. Stone, M.R. Hufford, Ecological momentary assessment, *Annual Review Clinical Psychology* 4:1 (2008):1–32.
- [30] E.O. Lillie, B. Patay, J. Diamant, B. Issell, E.J. Topol, and N.J. Schork, The n-of-1 clinical trial: the ultimate strategy for individualizing medicine? *Personalized Medicine* 8(2) (2011), 161–173.
- [31] R.L. Kravitz, N. Duan, eds, and the DECIDE Methods Center N-of-1 Guidance Panel, Design and Implementation of N-of-1 Trials: A User’s Guide, *AHRQ Publication*, 2014.
- [32] Kluge HE. The IMIA Code of Ethics for Health Information Professionals. IMIA endorsed document 2016-08-28. <http://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-Code-of-Ethics-2016.pdf> [last visit 2017-03-30]

Address for correspondence

Corresponding author: Pei-Yun Sabrina Hsueh

Email: pyhsueh@gmail.com

Extraction and Quantification Clusters of Three-Dimensional Lorenz Plots

Min Hu^a, Weixing Han^b, Jun Liu^c, Chengfan Jang^a, Suxia Wang^a

^a Department of Electrocardiogram, the People's Hospital of Huangshan, Huangshan, Anhui, People's Republic of China

^b Department of Electrocardiogram and Cardiac-function, the First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, People's Republic of China

^c School of General Education, Anhui Medical University, Hefei, Anhui, People's Republic of China

Abstract

Lorenz plot (LP) method gives a global view of long-time electrocardiogram signals, is an efficient simple visualization tool to analyze heart rate variability and cardiac arrhythmias. 50 records of RR-interval time series with frequent premature complexes exported from 24-hour Holter. Constructed three-dimensional LP (3DLP) with three successive RR intervals as X, Y and Z axis in Cartesian coordinate system for each record, then stereographic projection it along the space diagonal. The radii of dots distinguished eccentric clusters (ECs) points from centric clusters' with the accuracy of $94 \pm 6.0\%$. The eccentric scatter-dots were separated and identified by the frequency distribution characteristic on azimuth, with the accuracy of $93 \pm 13.3\%$ ECs. The $A_{PF} < -2.8^\circ$ of CPN EC supported the ventricular extrasystoles diagnosis, with excellent sensitivity (1.0) and specificity (0.92). The transformed coordinates (polar radius and angle) of 3DLP could extract and quantify clusters to diagnose arrhythmia, and might provide additional prognosis information.

Keywords:

Arrhythmia; Electrocardiography; Prognosis

Introduction

Lorenz (Poincaré) plot (LP) is a scatterplot on which each RR-interval (RRI) is plotted against next RRI, and it has been confirmed as a simple visualization tool to analyze heart rate variability (HRV) and cardiac arrhythmias [1]. Esperer reported that the LP method had the potential to significantly improve the accuracy of arrhythmia detection and differentiation [2].

In fact, this method highly depends on the recognition and classification of distinct scatter distribution patterns, i.e. the numbers and the relative positions of the scattered dots gathering regions, which are also known as clusters or attractors [1]. However, there is no efficient method of automatic recognition of LP morphologies, not to mention the method of cluster extraction and quantification [3]. So far, doctors still examine and analyze cases by the eyeball method, which is limited by their subjective experience. On the other hand, investigators might be interested in the number, morphology, position, size or density of clusters, because these may reflect the underlying information of corresponding cardiovascular status and cardiac electrophysiology. Some indices of LP have been extracted from clusters characteristics to evaluate the risks

of occurrence of lethal cardiac arrhythmia [4], or predict the probability of onset or termination of paroxysmal arrhythmia [5]. Then again, HRV has become an important method for assessing cardiovascular autonomic regulation. Most quantitative markers may be influenced by noise, ectopic beats, or other rhythm abnormalities, especially in frequent condition [6]. But, non-stationarities are important both as confounders and information conveyors. We assumed the quantification indices of clusters may supply additional information of predicting the occurrences and outcomes of arrhythmias and other cardiovascular diseases.

Methods

3DLP construction and clusters extraction

The RRI series were described as a time series vector $[x_1, x_2, \dots, x_n]^T$. The 3DLP was constructed of the plot of x_{i-1} on the x-axis, x_i on the y-axis and x_{i+1} on the z-axis in three-dimensional Cartesian coordinate system, that was, the (x_{i-1}, x_i, x_{i+1}) . Stereographic projection (SGP) of 3DLP along the space diagonal, there were one central cluster (CC) and four or more eccentric clusters (ECs) (see Figure 1.a and b). The clusters or points were located by polar coordinate system's radial (r_{2d}) and angular (t_{2d}) coordinates while the pole was set at $(45^\circ, 35.3^\circ)$. Divided the polar angular into 361 areas from $-\pi$ to π then counted the eccentric points of each area. The histogram of each EC presented normal distribution characteristic, with a distinct peak frequency in the middle (A_{PF}), and absolutely low frequency clockwise (A_{RF}) and counterclockwise (A_{LF}) borders (Figure 2). A searching procedure was designed to acquire such azimuths of each wave until the A_{PF} was small enough.

Clusters' extractions were automatic segmentations of 3DLP based upon homogeneously statistical properties of each coordinate element. The r_{2d} were used to distinguish CC points from ECs', and the azimuths (A_{PF} , etc.) were used to separate and identify ECs. The spherical radius (r_{3d}) of centric cluster permit to depart each part of it, if exist.

Professional software was developed to construct and extract clusters by our study group based on Matlab R2013. The development and running platform was Windows 7 operating system shipped on a personal computer, CPU 3.3 GHz, memory 2.0 GB.

Data source

The 50 cases of 24-hour Holter recordings obtained from our laboratory of the People's Hospital of Huangshan, which were artifact-free recording time of at least 22 hours and only frequent (>1000 times) unifocal (atrial or ventricular) premature complexes without other arrhythmias. The apparatus was DMS 300-3 Holter recorder and analyzer system, 128 Hz sample frequency. All records had been verified carefully by experienced cardiologist. Redefined the RRI annotations that: a normal-normal RRI as N, a pre-ectopic coupling interval as C, and a post-ectopic pause interval as P, the intervals preceded or followed an artifact, infrequent focal or interpolated extrasystole were marked as Z. The 3D dot-marker was composed of its successive three RRIs annotations (such as NPC) and used as standard classification of dots.

Statistical analysis

Quantitative results were indicated as mean values with pertinent standard deviations, or as percentages. The receiver operating characteristic (ROC) method was used on each record to explore the optimal threshold values, with the areas under the curves (AUC) for each feature represented by the ROC area.

Results

Frequent atrial or ventricular premature complexes records had 25 apiece. In each record, the numbers of RRIs were more than 80,000. The amount of P labelled RRIs was more than 1000; the total of Z labels was less than 100. Only clusters with the amounts of points more than 100 were accepted to analyze.

3DLP displayed multi-distribution subgraphs solid pattern emitted from the origin, more or less (Fig.1.a). Each cluster had unique color after colorized dots of each marker with a kind of color. That means a cluster was composed of points which had same pathological mechanism. By using the 3D dot-markers, it's easy to distinguish CCs from ECs for the former was composed of three exactly same RRI annotations, but the latter otherwise. It's also efficient to differentiate each EC.

The pathological mechanism of ECs was a regular ratio of sinus and ectopic beat occurrence. Each mark's EC had a relatively consistent A_{PF} between different patients and between different types of arrhythmias. They could be NNC, NCP, CPN and PNN when single premature beat occurred (A_{PF} located about at $-91\pm10.7^\circ$, $135\pm4.9^\circ$, $-2\pm7.8^\circ$ and $-151\pm2.6^\circ$ on SGP, respectively), CPC and PCP happened on bigeminy (about at $-30\pm12.0^\circ$ and $159\pm5.4^\circ$, respectively), PNC arose in trigeminy (about at $-119\pm7.3^\circ$), PCC appeared in true trigeminy condition (about at $-148\pm5.7^\circ$), NCC and CCP took place in the situation of paired or salvo premature beat onset or ending (about at $-142\pm6.1^\circ$ and $90\pm14.9^\circ$, respectively). The CC common was NNN cluster sourced from normal sinus rhythm, occasionally, accompanied with CCC if existed any salvo premature beats.

All 50 records, the mean numbers of 3D points were 94774 ± 11756 . The ROC method found that the mean cutoff value for r_{2d} was $4.7\pm2.83^\circ$ led the mean of sensitivities 0.95 ± 0.07 and specificities 0.91 ± 0.13 to distinguish both with AUC 0.95 ± 0.09 . Once adopted 4.7° as the cutoff value, the mean numbers of correct classified points were 89651 ± 13813 , accuracy rate $94\pm6.0\%$. Besides, we deduced that the minimal r_{2d} value of the lowest frequency of the histogram of r_{2d} could be used to identify sinus arrhythmia and atrial fibrillation,

because the former was less often than 10° and the latter usually greater than 17° .

The mean thresholds of r_{3d} from 9 records which met the statistical requirement was 1166 ± 189.3 ms lead the mean of sensitivities 0.87 ± 0.10 and specificities 0.92 ± 0.09 to separate the CCC from NNN cluster with AUC 0.93 ± 0.08 . The numbers of CC points were 45724 ± 29359 , once adopted 1166 as the cutoff value, correct classified 42628 ± 28543 ($89.8\pm7.51\%$).

Adjacent clusters had distinct A_{PF} values (the difference was more than 10° except PNN, PCC and NCC). It's possible to recognize the ECs by the A_{PF} values generated from the extraction procedure when RRIs annotations were not available, i.e., calculate the absolute values of the acquired A_{PFk} minus the experienced A_{PFS} (as described above), if someone (let NCP) was the minimum one and the difference was less 10° , then the k th cluster was that labelled clusters (i.e. NCP, and so on). There were 6 ± 2.6 ECs in the 50 cases; the numbers of correct recognition were 5 ± 1.6 , accuracy rate $93\pm13.3\%$. The numbers of ECs points were 27452 ± 27790 , correct classified 23822 ± 23930 ($89\pm12.6\%$).

The A_{PF} is a powerful index of differential diagnosis of atrial and ventricular extrasystoles. The ROC method found that the A_{PF} of CPN, PNC and NCP classified ventricular well from atrial, yielded AUC 0.98, 0.93 and 0.91 (all $P<0.01$), respectively. The $A_{PF} < -2.8^\circ$ of CPN cluster supported the ventricular extrasystoles diagnosis, with excellent sensitivity and specificity of 1.0 and 0.92, respectively. The result showed that 96% (48 of 50) of cases classified correctly.

Discussion

Extraction and quantification clusters of 3DLP

The quantification indices of LP clusters may supply additional information of predicting the occurrences and outcomes of arrhythmias and other cardiovascular diseases [2]. It's necessary to extract each cluster first. The dot-markers composed of RRI-annotations are helpful to do so, and make an opportunity to study the pathological mechanism of clusters. But they are not always available in clinical practice. It's necessary to develop an efficient method to achieve the goals.

In this paper we constructed a kind of 3DLP, in which, 3D scatter point was plotted with three successive RRIs as X, Y and Z axis in Cartesian coordinate system [1]. The clusters' space contours are more distinct for less overlap each other. The polar radius and angle of stereographic projection 3DLP along the space diagonal allow separation and recognition centric cluster and each eccentric cluster, and the spherical radius of centric cluster permit to depart each part of it. Under the help of dot-markers, the optimal threshold values were extracted. The results show that the proposed method has excellent accuracy of extraction and recognition clusters. The indices used to extract clusters may also be adopted to quantify the clusters, such as the polar radius and angle, and the spherical radius. The peak value or the up and down limit of such parameters refers to the central and dispersion tendency of a cluster. But it's still uncertain what additional cardiovascular pathology or physiology information could be provided by such parameters [1].

Relationships between 3DLP and 2D LPs

The morphologies of 3DLP viewed on XY or YZ plane are identical to the first-order of LP in every detail (Figure 3.a). The stereographic projection of 3DLP and the second-order of LP are much more alike than different (Figure 3.b). They are similar in amount and pattern of clusters, but have minor difference in radius, azimuth and overlap portions. Once we add a fourth dimension, the time dimension, which may be hidden, to 3DLP, it's easy to realize the retrograde technique, i.e., inspect any points on tachograms, even the ECG waves of the corresponding RRs, and vice versa, since both of them have a unique time dimension. Different clusters of 3DLP belong to different layers on tachograms when selected Y coordinates of 3D points correspond to y-axis of tachograms (Figure 4). Since the strong relationship exists among those different LPs, it indicates that we can apply the extracted clusters to conventional 2D LPs, and the majority of 2D LPs achievements could be utilized to 3DLP as well.

Limitations

We focused on extraction clusters of frequent arrhythmias in 3DLPs, which implied that arrhythmic episodes with a low incidence were ignored. There were few clusters still impossible to be extracted by this method, for example, if the PNN, PCC and NCC coexist. Atrial and ventricular arrhythmias coexistence in a 24-hour Holter recording was common; this study provided no methods to separate each other due partly to LPs lose QRS morphological information. This study didn't inspect all types of arrhythmias for insufficient data, nevertheless the proposed methodology provided the basic principles of how clusters are assessed.

Conclusion

This study provided two methods to extract clusters of 3DLP. The first one is gathering same markers' dots as a cluster, where the marker is composed of annotations of successive 3 RRs which constructed this dot. This method requires high quality of ECG recordings and the recordings has been analyzed accurately. The second one is separating dots into different clusters by its frequency distribution characteristic on space coordinates. This method is applicable on raw RRs data but may generate some deviation.

3DLP has significant advantage of extraction and recognition clusters for sufficiently less overlaps than 2D LPs. It's also benefited from the strong relationships with multiple 2D LPs.

Figures and Graphs

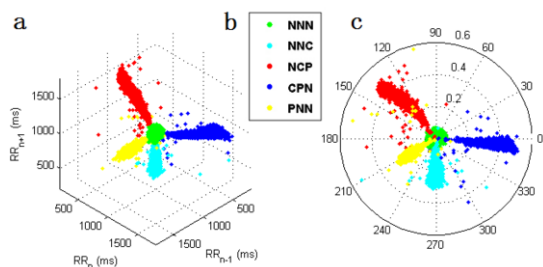


Figure 1 – (a) 3DLP (viewpoint at azimuth = 135° and elevation = 35.3°); (b) the colormap; (c) Stereographic projection of 3DLP along the space diagonal.

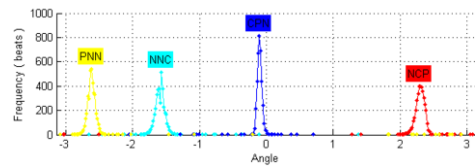


Figure 2 – Azimuth-frequency histogram of each ECs

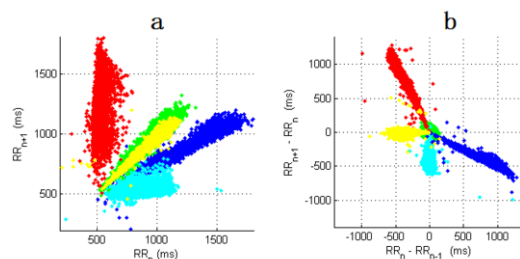


Figure 3 – First-order (a) and second-order (b) LP

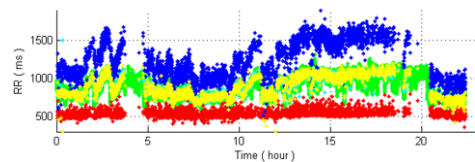


Figure 4 – The layers of clusters on tachogram

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] R. Sassi, S. Cerutti, F. Lombardi, M. Malik, H.V. Huikuri, C.-K. Peng, G. Schmidt, Y. Yamamoto, B. Gorenek, G.Y.H. Lip, G. Grassi, G. Kudaiberdieva, J.P. Fisher, M. Zabel, and R. Macfadyen, Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society, *Europace* (2015).

- [2] H.D. Esperer, E. Chris, and C.R.J, Cardiac arrhythmias imprint specific signatures on Lorenz plots, *Annals of Noninvasive Electrocardiology* **13** (2008), 44-60.
- [3] L. Zhang, T. Guo, B. Xi, Y. Fan, K. Wang, J. Bi, and Y. Wang, Automatic recognition of cardiac arrhythmias based on the geometric patterns of Poincare plots, *Physiol Meas* **36** (2015), 283-301.
- [4] J.P. Kelwade and S.S. Salankar, Prediction of Cardiac Arrhythmia using Artificial Neural Network, *International Journal of Computer Applications* **115** (2015), 30-35.
- [5] H. Lu, C. Zhang, Y. Sun, Z. Hao, C. Wang, and J. Tian, [Primary Study on **Predicting the Termination of Paroxysmal** Atrial Fibrillation Based on a Novel RdR RR Intervals Scatter Plot], *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* **32** (2015), 763-766.
- [6] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, Heart rate variability. Standards of measurement, physiological interpretation, and clinical use, *Eur Heart J* **17** (1996), 354-381.

Address for correspondence

Corresponding author name: Min Hu.

Please contact us on email.

Address: No. 4 Li Yuan Road, Huangshan, Anhui, China.

E-mail address: humin_jx@163.com, humin_jx@aliyun.com

Phone: +8605592521226; 13965520499.

E-Health Literacy and Health Information Seeking Behavior Among University Students in Bangladesh

Md.Mohaimenul Islam^{a,b}, Musa Touray^a, Hsuan-Chia Yang^b, Tahmina Nasrin Poly^c,
Phung-Anh Nguyen^b, Yu-Chuan (Jack) Li^{a,b}, Shabbir Syed Abdul^{a,b}

^aDepartment of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan, China

^bInternational Center for Health Information Technology (ICHIT), Taipei Medical University, Taiwan, China

^cDepartment of Microbiology and Immunology, Tzu Chi University, Hualien, Taiwan, China

Abstract

Web 2.0 has become a leading health communication platform and will continue to attract young users; therefore, the objective of this study was to understand the impact of Web 2.0 on health information seeking behavior among university students in Bangladesh. A random sample of adults (n= 199, mean 23.75 years, SD 2.87) participated in a cross-sectional, a survey that included the eHealth literacy scale (eHEALS) assessed use of Web 2.0 for health information. Collected data were analyzed using a descriptive statistical method and t-tests. Finally logistic regression analyses were conducted to determine associations between sociodemographic, social determinants, and use of Web 2.0 for seeking and sharing health information. Almost 74% of older Web 2.0 users (147/199, 73.9%) reported using popular Web 2.0 websites, such as Facebook and Twitter, to find and share health information. Current study support that current Web-based health information seeking and sharing behaviors influence health-related decision making.

Keywords:

Health Literacy; Social Media; Cross-Sectional Studies

Introduction

Nowadays, the Internet is a very common tools to seek information about healthcare and health conditions [1]. A study has confirmed that 83.4% of the frequent internet users age between twenty years to forty years, shows 72% of them are engaged in social networking days and nights [2]. As a large amount of health information is available in web 2.0; so it can be used to educate and empower people. The concept of eHealth literacy which refers to the ability to read, understand and communicate about health information to make the people to take proper health decision [3].

Consumer-directed eHealth requires the ability to seek out, find, evaluate and appraise, integrate, and apply what is gained in electronic environments toward solving a health problem, or eHealth literacy [4]. This composite skill requires that people are able to work with technology, critically think about issues of media and science, and navigate through a vast array of information tools and sources to acquire the information necessary to make decisions [3]. According to Norman and Skinner [5], eHealth literacy is ability to navigate the internet for health information. Thus, eHealth literacy comprise of

computer literacy, scientific literacy, health literacy, traditional literacy, media literacy, and information literacy [6]. Norman and Skinner [7] created the eHealth Literacy Scale (eHEALS) to measure individuals' perceptions of their own digital health literacy skills [8].

Although many studies have used eHEALS scale to determine eHealth literacy despite of lack of evidence [9-13]. Asian countries like Japan and Taiwan researcher used this eHEALS scale for predicting eHealth literacy. But there is no research in Southeast Asian countries using eHEALS scale to determine eHealth literacy. In this study, we construct validity of eHEALS was analyzed among the university students who use web 2.0.

Methods

Study population

A cross-sectional study was conducted in four university (Dhaka Dental College, American International University Bangladesh, Manarat University Bangladesh, and Stamford University Bangladesh). Participants were eligible if they were over 17 and less than 35 years old, capable of reading, writing English, and were willing to sign a consent form and able to complete questionnaire. A total of 199 participants were enrolled 4 investigators from November 2015 to March 2016. All of the 199 participants in this study were included according to gender, age, marital status, computer competency, and pattern of internet use.

Measurement

Computer knowledge and Internet use

Participant were asked about their computer knowledge and amount of internet use. In the past 12 months have they use popular social media and shared any information regarding health. It was not restriction to use desktop or laptop computer, cell phone, mobile handheld device like an e-reader or tablet.

eHealth literacy

In our study, we measured eHealth literacy by using eHealth literacy scale (eHEALS). It is introduced by Norman which is used determines consumers' combined knowledge, confidence, and perceived skills finding, evaluating, and applying electronic health information to health problems. eHEALS consists of 8-items scored on a 5-point Likert scale ranging

from 1 (strongly disagree) to 5 (strongly agree). Higher scores on the eHEALS indicates higher eHealth literacy (total score range=5-40). The internal consistency of the data collected using the eHEALS in this study was high (Cronbach alpha=.740), and comparable to reliability estimates reported in previous studies.

Use of Web 2.0 for Health Information

We asked participants, “In last 12 months, have you used the Internet for any of the following reasons to locate or share health information?” Respondents could select all reasons for using the Internet:

1. Participated in a Web-based-support group,
2. Used a social networking site like Facebook/Twitter/LinkedIn,
3. Wrote in a Web-based diary or blog.

Sociodemographic and Social Determinant Variables

Gender (male or female), age (in years), education (higher secondary school/H.S.C, college graduate, post-graduate), and marital status (married, unmarried, widow and unknown) were all assessed. Perceived health status was also measured using the following scale: (1) poor, (2) fair, (3) good, (4) very good, and (5) excellent.

Statistical Analysis

In our study, we used SPSS version 23.0 to compute frequency and descriptive statistics to analyses sociodemographic and social determinant characteristics, frequency statistics for each eHEALS item, and the number of respondents reporting use of Web 2.0 for health information. An independent samples t-test was performed to compare eHealth literacy among users and non-users of Web 2.0 for health information. We also conducted a multiple linear regression to determine whether use of internet and computer knowledge, sociodemographic variables (sex, age, education, marital status), and perceived health status as a determinant predicted overall eHEALS scores. Finally, a binominal logistic regression was conducted to determine whether these predictor variables were associated with the use/non-use of Web 2.0 for health information. Analyses were considered statistically significant at the $P<0.05$ alpha level (two-tailed).

Results

Participant characteristics

Our study shows that participants age ranged from 17 to 35 years (mean 23.75, SD= 2.87). Male participant had 133 (66.8%) and female had 66 (33.2%). Among participant unmarried was highest number 138 (69.3%), 48 (24.1) were married, 1 (0.5) was widow and 12 (6%) did not answer this question. Over 72% of participant reported completing Bachelor degree and almost a quarter 23.1 % (46/199) completed Master's degree. A little over one quarter (27.6%) participants had very good health and 30.7% (61/199) participants had good health status. **Table 1:** shows the characteristics of study participants included this study (n=199).

Computer competency and internet use

A little over half of the participants were competent in computer 53.3% (106/199), 31.7% (63/199) of despondences were just beginner in computer and about one six percent 15.1% (30/199)

despondences were above competent. In the case of internet use, 82.9% (165/199) of participants use internet daily, 11.6% (23/199) use internet once a week.

Table 1 - Sociodemographic and health status, computer knowledge and internet use characteristics of study participants (N=199)

Demographic	N (%)
Gender	Male 133 (66.8) Female 66 (33.2)
Mean (SD) age (Years)	23.75 (2.87)
Marital status	Married 48 (24.1) Unmarried 138(69.3) Widow 1 (0.5) Unknown 12 (6)
Education	Higher secondary 9 (4.5) Bachelor 144 (72.4) Masters 46 (23.1)
Health status	Excellent 56 (28.1) Very good 55 (27.6) Good 61 (30.7) Fair 18 (9) Poor 2 (1) Not answer 7 (3.5)
Computer knowledge	Beginner 63 (31.7) Competent 106 (53.3) Above competent 30(15.1)
Use of Internet	Daily 165(82.9) Once a week 23 (11.6) More than one times a week 7 (3.5) Once a month 4 (2.0) Never 0 (0)

Use of Web 2.0 for health information

Almost three fourth of participants 71.9% (143/199) use popular social media like face book, twitter etc. and share health information. However, almost 40% (79/199) of participants does not share any information in web based support group and 94% (187/199) of participant does not share any information in blog. Almost three fourth 73.9% (147/199) of participant reported use at least one social media for sharing health information. **Table 2:** shows the frequency and percentage of young's who used Web 2.0 to locate or share health information.

Table 2 - Frequency and percentage of young's who used Web 2.0 to locate or share health information (N=199).

In last 12 months, have you used the Internet for any of the following reasons to locate or share health information?	N (%)
Popular social media	
Yes	143 (71.9)
No	56(28.1)
Web-based support group	
Yes	79(39.7)
No	120(60.3)
Blogs	
Yes	12(6)
No	187(94)
Report using at least one of these types of social media	
Yes	147(73.9)
No	52 (26.1)

Reliability and validity

In our study, total scale of eHEALS ranged from 13 to 40 (means 27.46, SD= 4.99). Table 3 illustrates the response frequencies for each eHEALS items. The internal consistency of the eHEALS was alpha= 0.74. The variance of the scale was 50.7 percent and all items loaded high on this component,

ranging from .423 to .642. **Table 3:** shows the correlation the scores on the eHEALS and the variable measured in this study.

Association between Web 2 for health information and eHealth literacy

In the case of social media use, the respondents reported the difference in the total eHEALS scores among popular social media such as Facebook, twitter, etc. users (means 28.01, SD= 4.95) and non-users (means 25.98, SD= 4.79). Other response in this question included the difference in the total eHEALS scores among the web support group users (means 27.75, SD= 4.28) and non-users (means 27.24, SD= 5.398). This difference is also observed in the case of blogs, the difference in the total eHEALS scores among blog users (means 27.83, SD= 5.95) and non-users (means 27.42, SD= 4.93). Taken together, these results suggest that there is an association between Web 2 for health information and eHealth literacy.

Predictors of Web 2 use for health information

Binominal logistic regression analysis were used to analyze the relationship between web 2.0 use and eHealth literacy. **Table 4** provides the summary statistic for predicting use of web 2 for the health information.

Discussion

The present study was designed to determine the effect of internet on health literacy among the university students in

Bangladesh. In this study sociodemographic variable such as health status and social determinant (e.g. marital status, income etc.) were not significant predictors of eHealth literacy among the university students in Bangladesh. However, gender appeared significant effect especially female significantly effects the use of internet on health information. This finding is contrary to previous study which has suggested that education level, advanced age, and the extent to which electronic devices were used did appear to affect eHealth literacy. Also, they found the level of education, electronic device use influenced the use of internet for health-related information [14]. Furthermore, this present study found that the majority of students used the Internet to find health information, and believed the Internet was useful for helping to make health decisions.

E-Health literacy was found to be influenced by age, education, and marital status, computer knowledge used to search for health information in this study. Participants having higher level of education has been associated with higher amount of internet use for searching health information. Although, Powel et al. [15] studies mentioned education has been associated with more frequent use of the Internet for health information. This view is also supported by a group of researchers from Israel [12]. Besides, unmarried participant's searching frequency were higher than married participants. In addition, computer competency had great influence in

Table 3 - eHEALS scale mean, SD, reliability and factor analysis

Items	Mean	SD	Factor loading	Item-total correlation
I know what health resources are available on the Internet	3.61	.874	.423	.370
I know where to find helpful health resources on the Internet	3.57	.950	.582	.486
I know how to find helpful health resources on the Internet	3.51	1.024	.642	.508
I know how to use the Internet to answer my health questions	3.32	1.171	.594	.470
I know how to use the health information I find on the Internet to help me	3.61	.880	.468	.395
I have the skills I need to evaluate the health resources I find on the Internet	3.30	1.158	.565	.465
I can tell high quality from low quality health resources on the Internet	3.24	1.159	.493	.407
I feel confident in using information from the Internet to make health decisions	3.30	1.114	.433	.389
Mean (SD) sum score	27.46	4.99		
Variance accounted for	50.07%			
Cronbach alpha	.740			

Table 4 - Logistic regression predicting use of Web 2.0 for health information

Variables	B	SE B	Wald	Exp (º)	95% CI
Constant					
Age	-.026	.063	.17	.974	.86-1.10
Gender	-.407	.368	1.22	.666	.32-1.37
Marital status ^b					
Married	1.887	1.145	2.71	6.601	.70-62.2
Unmarried	1.594	1.099	2.10	4.923	.57-42.4
Education ^c					
Bachelor	.310	.390	.63	1.363	.63-2.92
Master's	.532	.908	.34	.587	.09-3.48
Computer knowledge ^d					
Beginner	.169	.394	.18	1.184	.54-2.56
Competent	.286	.540	.28	1.332	.46-3.83
Internet use ^e					
Daily	1.163	.502	5.36	3.201 ^a	1.19-8.56
Once a week	.540	.892	.36	1.715	.29-9.86
More than one a week	1.116	1.067	1.29	3.374	.41-27.32

^aP<.05 two-tailed; ^bReference category: Not answer; ^cReference category: Higher secondary school certificate;

^dReference category: Above competent; ^eReference category: Once a month

searching health information in internet and it shifted from beginners to competent. Several studies mentioned that demographics, educational background, and technology use uniquely influences health literacy [16-18] and eHealth literacy [4; 19; 20] in the general population.

As Bangladesh is a developing country so there has been some concern that those at lower socio-economic levels do not have equal access to Internet health resources. It is clearly understandable that access to computers are limited, literacy abilities are insufficient, and always lacking of basic computer skills. However, it is true that university students are still the largest percentage of Internet health information seekers and it also exponentially increasing to the other segments of society, obviously they are likely to search for health information. As significant number of university student are using Internet for seeking health resources but how they look for and find high-quality information on the Web still not clear.

While it is important to use the Internet to seek out general health information, majority of adult does not feel interest to discuss their own health problems or obtain personalized medical advice over the Internet. Most of cases they are reluctance to using interactive Internet applications for health communication purposes. It would be more valuable if they seek and share their personal health information other than just usual interaction. It could be the result of contextual Web security issues affecting confidentiality. The issue of trust when using the Internet to seek and share medical information is an important one to consider, especially with the emergence of peer-to-peer or horizontal health communication among university students. More research should be done to discover what particular sources of Web-based health information college students are consulting and which cause uneasy feelings originating from potential threats to data security and privacy.

There are some limitation in our study. Firstly, we only focus young generation who are studying different university; we did not include participants who were not studied in institution. Secondly, we only conducted survey one district, it does not present whole country situation. There may be a need for a more comprehensive survey instrument that assesses health information seeking and sharing using all types of Internet applications. Finally, the cross-sectional research design limits the researchers from establishing causation when considering the interrelationships between sociodemographic variables, social determinants, and health communication outcomes.

Conclusion

Although university students are highly connected to, and feel comfortable with, using the Internet to find health information but it is not still satisfactory. This study has shown that eHealth literacy enhancing program is needed among university students. It is important those who are in the medical and health professions, need customized eHealth literacy training for finding, interpreting, and evaluating health- and medical-related information available on the Internet.

References

- [1] C. Norman, eHealth literacy 2.0: problems and opportunities with an evolving concept, *Journal of medical Internet research* **13** (2011), e125.
- [2] C. Knapp, V. Madden, H. Wang, P. Sloyer, and E. Shenkman, Internet use and eHealth literacy of low-income parents whose children have special health care needs, *Journal of medical Internet research* **13** (2011), e75.
- [3] D. Cameron and A. Harvey, eHEALS: The eHealth Literacy Scale, *J Med Internet Res* **8** (2006), e27.

- [4] C.D. Norman and H.A. Skinner, eHealth literacy: essential skills for consumer health in a networked world, *J Med Internet Res* **8** (2006), e9.
- [5] C. Robinson and J. Graham, Perceived Internet health literacy of HIV-positive people through the provision of a computer and Internet health education intervention, *Health Information & Libraries Journal* **27** (2010), 295-303.
- [6] J. Nguyen, M. Moorhouse, B. Curbow, J. Christie, K. Walsh-Childers, and S. Islam, Construct Validity of the eHealth Literacy Scale (eHEALS) Among Two Adult Populations: A Rasch Analysis, *JMIR public health and surveillance* **2** (2016), e24.
- [7] C.D. Norman and H.A. Skinner, eHEALS: the eHealth literacy scale, *Journal of medical Internet research* **8** (2006), e27.
- [8] R.R. Sinkovics, T. Salzberger, T. Salzberger, and R.R. Sinkovics, Reconsidering the problem of data equivalence in international marketing research: Contrasting approaches based on CFA and the Rasch model for measurement, *International Marketing Review* **23** (2006), 390-417.
- [9] W. Hsu, C. Chiang, and S. Yang, The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy, *Journal of medical Internet research* **16** (2014).
- [10] S. Mitsuake, A. Shibata, K. Ishii, K. Okazaki, and K. Oka, [Developing Japanese version of the eHealth Literacy Scale (eHEALS)], [*Nihon koshu eisei zasshi*] *Japanese journal of public health* **58** (2011), 361-371.
- [11] B. Xie, Effects of an eHealth literacy intervention for older adults, *Journal of medical Internet research* **13** (2011), e90.
- [12] E. Neter and E. Brainin, eHealth literacy: extending the digital divide to the realm of health information, *Journal of medical Internet research* **14** (2012), e19.
- [13] C.A. Brown and R. Dickson, Healthcare students'e-literacy skills, *Journal of allied health* **39** (2010), 179-184.
- [14] B. Tennant, M. Stelfox, V. Dodd, B. Chaney, D. Chaney, S. Paige, and J. Alber, eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults, *Journal of medical Internet research* **17** (2015), e70.
- [15] J. Powell, N. Inglis, J. Ronnie, and S. Large, The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study, *Journal of medical Internet research* **13** (2011), e20.
- [16] L. Nielsen-Bohlman, A. Panzer, B. Hamlin, and D. Kindig, Institute of Medicine. Health literacy: a prescription to end confusion. Committee on Health Literacy, Board on Neuroscience and Behavioral Health, in, Washington, DC: National Academies Press, 2004.
- [17] D. Nutbeam, Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century, *Health promotion international* **15** (2000), 259-267.
- [18] R.L. Rothman, D.A. DeWalt, R. Malone, B. Bryant, A. Shintani, B. Crigler, M. Weinberger, and M. Pignone, Influence of patient literacy on the effectiveness of a primary care-based diabetes disease management program, *Jama* **292** (2004), 1711-1716.
- [19] R. Van der Vaart, A.J. van Deursen, C.H. Drossaert, E. Taal, J.A. van Dijk, and M.A. van de Laar, Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations, *Journal of medical Internet research* **13** (2011), e86.
- [20] N.G. Choi and D.M. DiNitto, The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use, *Journal of medical Internet research* **15** (2013), e93.

Address for correspondence

Dr. Shabbir Sayed Abdul MD, MSc, PhD
Assistant Professor and Health IT consultant
Graduate Institute of Biomedical Informatics,
College of Medical Science and Technology
Taipei Medical University, Taiwan

Comparing Cancer Information Needs for Consumers in the US and China

Zongcheng Ji^a, Yaoyun Zhang^a, Jun Xu^a, Xiaoling Chen^a, Yonghui Wu^a, Hua Xu^a

^a School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, USA

Abstract

Due to the differences in environments and cultures, consumers seeking cancer information in various regions of the world may have diverse needs. This study compares the cancer information needs for consumers in the US and China. Specifically, we first collected 1,000 cancer-related questions from Yahoo! Answers and Baidu Zhidao, respectively. Then, we developed a taxonomy of health information needs and manually classified the questions using the taxonomy. Finally, we analyzed the characteristics of information needs from consumers in both countries and summarized the differences between them. Our study demonstrated that although there are some common needs between consumers in the US and China, there are several significant differences between the two countries: the Chinese consumers are more likely to seek diagnosis and treatment online, while the US consumers prefer to seek common medical knowledge online.

Keywords:

Cancer; Information Needs; Taxonomy; Consumer Health Informatics

Introduction

Cancer is one of the leading causes of death worldwide. Tremendous efforts from the government, health providers, and pharmaceutical companies have been devoted to developing effective cancer therapies. With the initiative of precision medicine, a recent trend in cancer treatment is to provide personalized cancer therapy [1; 2], in which the proactive engagement from patients regarding treatment preference and healthcare data sharing plays a critical role. Therefore, multiple healthcare organizations (e.g., Mayo Clinic¹) and government institutes (e.g., National Cancer Institute of US²) provide comprehensive online cancer knowledge services and have developed patient portals to answer healthcare questions and collect feedbacks, for better patient education and engagement in personalized cancer therapy.

Meanwhile, consumers (e.g., patients, family members, friends) seeking cancer information all over the world are also eager to know different cancer-related information, ranging from common knowledge about the causes, symptoms, and treatments to the most suitable healthcare providers and emerging new treatment options in the era of precision medicine. Moreover, people in different regions are attempting to find cancer knowledge and therapies worldwide to leverage better healthcare resources. Therefore, it is necessary to build automated informatics tools such as search engine or question answering systems that can integrate the rich resources and provide more efficient and effective cancer

knowledge services for consumers. One essential step toward this goal is to first understand the specific cancer information needs of consumers, in order to provide the most helpful information.

Fortunately, such information needs are expressed in various social communities, such as online question answering communities where people can ask their questions and answer others' questions (e.g., Yahoo! Answers³ in the US and Baidu Zhidao⁴ in China). As time passed, these websites have accumulated a large amount of cancer-related questions and answers, which provide us invaluable data resources for understanding consumer health information needs. However, people from different regions with different cancer distributions, therapy development and cultures may have different information needs [3; 4]. For example, our preliminary study reveals that consumers in the US prefer to know the causes or risk factors of cancers, while consumers in China pay more attention to seek possible treatments online. Identifying the differences in information needs among consumers in different regions will facilitate the building of more advanced automated cancer knowledge services (e.g., search engines / question answering systems) which can provide knowledge / answers according to consumer-centered information needs. Furthermore, being aware of the different information needs required by the consumers will also facilitate cross-region collaborations for cancer diagnosis and treatment, such as the rapidly emerging international telemedicine.

There are several studies [5-11] that have analyzed cancer-related questions to understand consumer information needs. These studies mainly focused on very specific topics, including the information needs of post-treatment cancer patients [7; 9], patients in self-characterized illness phase [8], breast cancer patients [10], rare cancer patients [6], melanoma patients [11] and cancer patients' questions about pain [5]. However, no comprehensive study of consumer information needs of cancers has been conducted. Furthermore, the differences in information needs among consumers in different regions of the world have not been previously investigated. This study compares the cancer information needs for consumers in different regions. As a starting point, we set the comparison between the US and China, because

¹ <http://www.mayoclinic.org/>

² <https://www.cancer.gov/>

³ <https://answers.yahoo.com/>

⁴ <https://zhidao.baidu.com/>

Table 1 – The detailed definitions for the 10 main categories of the taxonomy for consumer information needs related to cancers

Main Category	Description
<i>Common Knowledge</i>	Common medical knowledge, such as the definition, prevalence, etiology, prognosis or information resources of some conditions, etc.
<i>Diagnosis</i>	What are the possible conditions given certain symptoms, what are the possible symptoms given certain conditions, and knowledge about the tests such as the skin test, imaging and biopsy, etc.
<i>Treatment</i>	Treatments of diseases, which include the drug therapy, surgery, radiotherapy and targeted therapy, etc.
<i>Prevention</i>	The prevention of some cancers with drug, food or other methods.
<i>Healthy Lifestyle</i>	Healthy diet, mood control or other issues about how to keep healthy or help for recovery in daily life.
<i>Health Provider Choosing</i>	Information about choosing hospitals, departments, doctors or other healthcare organizations such as the support groups, associations and foundations, etc.
<i>Second Opinion</i>	Seek second opinions or suggestions from others.
<i>Similar Experience Finding</i>	Find someone who has similar experiences.
<i>Finance</i>	Information about financial support, such as the insurance, claim and charity, etc.
<i>Other</i>	Other questions that cannot fall into any categories above.

both countries have large scales of patients with different cancer distributions and cultures [3; 4]. Specifically, we first collected 1,000 cancer-related questions from Yahoo! Answers and Baidu Zhidao, respectively. Then, we developed a taxonomy of health information needs and manually classified the questions using the taxonomy. Finally, we analyzed the characteristics of information needs from consumers in both countries and summarized the differences between them. As far as we know, this is the first study to compare cancer information needs for consumers in the US and China. Our study will greatly benefit the development of automated cancer knowledge services for consumers as well as the cross-region collaboration for cancer therapy.

Methods

Datasets

We collected cancer-related questions from Yahoo! Answers and Baidu Zhidao, which represent the most popular online question answering communities in the US and China, respectively [12-16]. More specifically, we collected questions under the category of “Health / Diseases & Conditions / Cancer” from Yahoo! Answers and “医疗健康 / 肿瘤科 (Health Care / Oncology)” from Baidu Zhidao. To facilitate analysis, we took the *subject* field as the question without considering the *description* field, which contains the detailed description. Finally, we collected 9,043 and 166,469 questions in English and Chinese, respectively.

We randomly sampled 1,000 questions from each dataset to conduct the comparative study. As all the questions are user generated, some questions are not clearly described (e.g., Brain tumor?) or even not related to the cancer topic. We discarded these questions and randomly sampled new ones to keep the sample size at 1,000 for each dataset.

Taxonomy

To characterize consumer health information needs, a taxonomy is needed to classify the consumers' questions. The Taxonomy of Generic Clinical Questions (TGCQ) [17] is one of the most popular taxonomies for classifying health-related questions. Several studies [17; 18] show that TGCQ is useful for analyzing physicians' and case managers' information needs, but is not suitable for consumers' questions due to the difference in the information needs between the physicians and the consumers [19]. We are aware of one recent study [19] that developed a taxonomy mainly based on TGCQ for analyzing consumer health information needs on hypertension related questions. However, this taxonomy is incomplete or somewhat confusing for our case. For example, there is no category for the questions regarding seeking second opinions,

finding similar experiences and asking financial support, where these three categories take up 11.7% of the English questions in our study. Furthermore, we also created a new category named *Common Knowledge*, which covers the definition, prevalence, etiology, prognosis or information resources of some conditions, etc. Finally, based on TGCQ and the taxonomy developed in the work of [19], we developed our new taxonomy, which consists of two levels of categories. The first level contains 10 main categories and the second level contains 28 subcategories. The detailed definitions for the 10 main categories are shown in [Table 1](#) and the 28 subcategories can be found in [Table 2](#).

Annotation

We manually reviewed 100 questions from both the two datasets to develop an annotation guideline. [Table 3](#) shows several examples from the annotation guideline. For instance, the question “Why does cancer treatment work for some and not others??” can be classified into the main category Treatment, but cannot fall into any other subcategories of Treatment (i.e., Drug Therapy, Surgery, Other Treatments and Treatment Seeking). Thus, we classify this question as Treatment→Other. Two annotators who are fluent in both English and Chinese manually classified the 2,000 questions (1,000 in English and 1,000 in Chinese) with the taxonomy we developed. The disagreements in the annotations were solved through group discussion including the two annotators.

Statistical Analysis

Cohen's kappa [20] was used to calculate inter-annotator agreement scores for the first level categories and the second level categories. We examined the frequency distribution of cancer-related questions among the first level categories and the second level categories for both the English dataset and the Chinese dataset. We also compared the frequency distribution across the English questions and the Chinese questions to identify the differences.

Results and Discussion

Taxonomy reliability

The two annotators manually annotated the 1,000 cancer-related questions from Yahoo! Answers (US) and 1,000 cancer-related questions from Baidu Zhidao (China) using the developed taxonomy. [Table 2](#) shows the frequency distribution over the two-level categories for the two datasets. The kappa measurement implemented in Stata was used to calculate the inter-annotator agreement between the two annotators.

Table 2 – Two-level categories of consumer information needs and their frequencies on cancer-related questions

Main Category	Subcategory	Frequency on Yahoo Data	Frequency on Baidu Data
Common Knowledge	Definition	76 (7.6%)	41 (4.1%)
	Prevalence	73 (7.3%)	31 (3.1%)
	Etiology	115 (11.5%)	86 (8.6%)
	Prognosis	64 (6.4%)	43 (4.3%)
	Information Seeking	47 (4.7%)	7 (0.7%)
	Other	48 (4.8%)	36 (3.6%)
Diagnosis	Condition	56 (5.6%)	131 (13.1%)
	Symptom	48 (4.8%)	33 (3.3%)
	Test	35 (3.5%)	83 (8.3%)
	Other	75 (7.5%)	79 (7.9%)
Treatment	Drug Therapy	26 (2.6%)	66 (6.6%)
	Surgery	27 (2.7%)	28 (2.8%)
	Other Therapy	62 (6.2%)	157 (15.7%)
	Treatment Seeking	22 (2.2%)	22 (2.2%)
	Other	1 (0.1%)	1 (0.1%)
	Other	2 (0.2%)	6 (0.6%)
Prevention	Drug for prevention	9 (0.9%)	20 (2.0%)
	Food for prevention	6 (0.6%)	13 (1.3%)
	Other	18 (1.8%)	47 (4.7%)
Healthy Lifestyle	Diet	35 (3.5%)	5 (0.5%)
	Mood Control	11 (1.1%)	19 (1.9%)
	Other	17 (1.7%)	35 (3.5%)
Health Provider Choosing	Hospital	0 (0.0%)	6 (0.6%)
	Department	7 (0.7%)	7 (0.7%)
	Doctor	8 (0.8%)	0 (0.0%)
	Other	60 (6.0%)	7 (0.7%)
Second Opinion		33 (3.3%)	2 (0.2%)
Similar Experience Finding		24 (2.4%)	1 (0.1%)
Finance		33 (3.3%)	8 (0.8%)
Other			

Table 3 – Examples from the annotation guideline of consumer information needs related to cancers

Category	Guideline / Generic Types of Questions	Example Questions (English translations are given following the Chinese questions.)
Common Knowledge → Etiology	Causes such as genetic inheritance / risk factors of some conditions	Does salt lead to lung cancer? 血管瘤一定是先天性的吗? (Are all the cases of Hemangioma congenital?)
Treatment Treatment seeking	Questions about seeking help or treatment for some conditions.	How do you treat breast cancer? 平滑肌肉瘤如何治疗 (How to treat leiomyosarcoma)
Treatment → Other	Questions about treatment, which cannot fall into any other subcategories of Treatment.	Why does cancer treatment work for some and not others?? 淋巴瘤可以根治吗? (Is lymphoma curable?)
Second Opinion	Questions of seeking second opinions or suggestions.	If you have just had a very close neighbor die of liver cancer..what do you do? 去看望胃癌病人, 带什么比较好? (食物、水果) (To visit patients with gastric cancer, which is better to bring? Food or fruits?)
Similar Experience Finding	Questions of finding similar experiences.	Has anyone suffered from choriocarcinoma? 生命丝带对乳房腺瘤有用吗? 用过的朋友请进 (Is Life Ribbon useful for breast adenoma? Those who have used it, please come here)
Finance	Questions about financial support, such as the insurance, claim and charity, etc.	I would like to know if i can get affordable insurance for melanoma treatment? 军人得癌病, 需要终生服药, 复员时有相应政策吗? (Is there any demobilization policy for soldiers who have cancers and need life-long medication?)

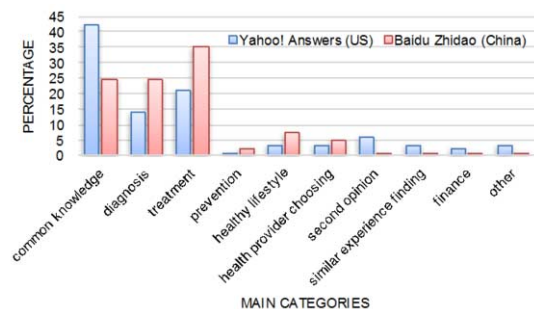


Figure 1 – Comparisons of frequency distributions over the 10 main categories for the questions from the US and China

For the first level categories, the two annotators achieved a Kappa score of 0.9105 and 0.9502 for the English questions and the Chinese questions, respectively. For the second level categories, the two annotators achieved a Kappa score of 0.8976 and 0.8798 for the English questions and Chinese questions, respectively. The good kappa scores indicated that the two annotators achieved high agreement on both of the two-level categories.

Frequency distribution of cancer information needs in US

Among the 1,000 English questions from Yahoo! Answers, the most frequent main category is *Common Knowledge* (42.3%), with the most frequent subcategory being *Etiology*, indicating that consumers in the US are more concerned with the cause or risk factors of diseases. The second most frequent main category is *Treatment* (21.2%), where the most frequent

subcategory is *Drug Therapy*. The third most frequent main category is *Diagnosis* (13.9%), where most of the questions are about the subcategory *Condition* (i.e., what are the possible conditions given the symptoms). For other main categories, the frequency distributions are 6% for *Second Opinion*, 3.5% for *Healthy Lifestyle*, 3.3% for *Similar Experience Finding*, 3.2% for *Healthcare Provider Choosing*, 2.4% for *Finance* and 0.9% for *Prevention*.

Frequency distribution of cancer information needs in China

Among the 1,000 Chinese questions from Baidu Zhidao, the most frequent main category is *Treatment* (35.2%), where the most frequent subcategory is *Treatment Seeking*, indicating that consumers in China care more about seeking possible treatments online. The second most frequent main category is *Diagnosis* (24.7%), where the most frequent subcategory is *Condition*. The third most frequent main category is *Common Knowledge* (24.4%), where the most frequent subcategory is *Etiology*. The frequency distribution for other main categories are 7.1% for *Healthy Lifestyle*, 4.8% for *Healthcare Provider Choosing* and 2% for *Prevention*. There are very few questions about requesting second opinions, finding similar experiences and asking for help with financial support.

Comparisons of cancer information needs between China and US

Figure 1 and Figure 2 shows the comparisons of frequency distributions over the 10 main categories and the 28 subcategories for the two datasets, respectively. From the two figures, we can find that the most frequent main categories for both datasets are *Common Knowledge*, *Diagnosis* and *Treatment*, which covers 77.4% and 84.3% of the questions in the two datasets, respectively. This indicates that consumers in the US and China have the similar concerns even with different healthcare systems and cultures.

Among the top three most frequent main categories of both datasets, we found several differences between the US and China. For the US consumers, the most frequent category is *Common Knowledge*, while it is *Treatment* for the Chinese consumers. This may indicate that the US consumers are more likely to ask about common medical knowledge, whereas the Chinese consumers are more likely to ask about the possible treatments for specific diseases. Among the subcategories of the top three frequent main categories, the Chinese consumers are more likely to ask about the *Treatment Seeking* and *Condition* (i.e., what are the possible conditions given the symptoms); whereas the US consumers are more likely to ask about the common knowledge related to *Etiology*. Why the Chinese consumers are more likely to seek diagnosis and therapies through online communities is not certain, but it is interesting and worth conducting further investigation.

Potential reasons may be the differences between the healthcare systems. For example, few Chinese have primary care providers, whereas many US patients can get useful information from their family doctors.

Beside the differences among the three common categories, we also found that the US consumers are more likely to ask questions about *Second Opinion*, *Similar Experience Finding* and *Finance* than the Chinese consumers (11.7% vs. 1%). This indicates that the US consumers are prone to exchange personal opinions and seek financial support, while the Chinese consumers not. This might be related to the differences between health insurance systems. Most of the Chinese consumers have government sponsored insurance, whereas the US patients are involved in very complex insurance systems. Again, further investigation should be taken to determine why this difference exists. On the other side, the Chinese consumers care more about the *Diet* for *Healthy Lifestyle* (4.7%), which indicates that more consumers in China consider the diet to be helpful for cancer treatment. This also reflects one of the Chinese traditional cultures – many Chinese consumers believe in food therapy.

Limitations

This study has several limitations. First, we limited the sources of English questions to Yahoo! Answers and Chinese questions to Baidu Zhidao. These two sources have different usability designs and demographic characteristics, which should be taken into account in our future work. In addition, questions from other online question answering communities will also be collected and analyzed in the next step. Second, we only focused on the general cancer domain and did not differentiate between different cancer subtypes. Our future work includes developing automated methods to mine information needs for different cancer subtypes. We are also planning to integrate this work into a medical question answering system to utilize the information needs embedded in the questions.

Conclusions

In this study, we identified the cancer information needs for consumers both in the US and China, and compared the differences. We developed a taxonomy of consumer health information needs regarding cancer and annotated two cancer-related datasets with the taxonomy. Based on the quantitative analysis, we reported some interesting observations. To the best of our knowledge, this is the first study focused on consumer information needs for cancer between the US and China. Our study will benefit the development of automated cancer-related search engine / question answering systems for consumers as well as the cross-region collaboration for cancer therapy.

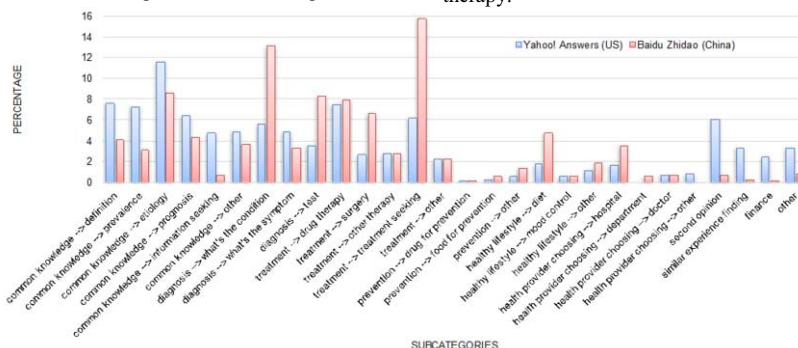


Figure 2 – Comparisons of frequency distributions over the 28 subcategories for the questions from the US and China

Acknowledgements

This study is supported in part by grants from NLM 2R01LM010681-05, NIGMS 1R01GM103859, and 1R01GM102282.

References

- [1] S. Heath, Facilitating Patient Participation in Precision Medicine, in: <https://goo.gl/dKcXDC>, 2016.
- [2] C. Petersen, The Future of Patient Engagement in the Governance of Shared Data, *eGEMs* **4** (2016).
- [3] W. Chen, R. Zheng, P.D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X.Q. Yu, and J. He, Cancer statistics in China, 2015, *CA: a cancer journal for clinicians* **66** (2016), 115-132.
- [4] R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2015, *CA: a cancer journal for clinicians* **65** (2015), 5-29.
- [5] J.L. Bender, J. Hohenadel, J. Wong, J. Katz, L.E. Ferris, C. Shobbrook, D. Warr, and A.R. Jadad, What patients with cancer want to know about pain: a qualitative study, *Journal of pain and symptom management* **35** (2008), 177-187.
- [6] R. Falotico, C. Liberati, and P. Zappa, Identifying Oncological Patient Information Needs to Improve e - Health Communication: a preliminary text - mining analysis, *Quality and Reliability Engineering International* **31** (2015), 1115-1126.
- [7] E.M. Galarce, S. Ramanadhan, J. Weeks, E.C. Schneider, S.W. Gray, and K. Viswanath, Class, race, ethnicity and information needs in post-treatment cancer patients, *Patient education and counseling* **85** (2011), 432-439.
- [8] J.E.Z.D.W. Pratt, Self-characterized illness phase and information needs of participants in an online cancer forum, (2015).
- [9] M. Shea-Budgell, X. Kostaras, K. Myhill, and N. Hagen, Information needs and sources of information for patients during cancer follow-up, *Current Oncology* **21** (2014), 165.
- [10] Y. Zhang and H. Xu, Understanding consumers ' information needs about breast cancer by analyzing online questions, in: *AMIA*, 2013, p. 1578.
- [11] A. Molassiotis, L. Brunton, J. Hodgetts, A.C. Green, V.L. Beesley, C. Mulatero, J.A. Newton-Bishop, and P. Lorigan, Prevalence and correlates of unmet supportive care needs in patients with resected invasive cutaneous melanoma, *Annals of Oncology* **25** (2014), 2052-2058.
- [12] D. Wu and D. He, A study on Q&A services between community-based question answering and collaborative digital reference in two languages, *IAENG International Journal of Computer Science* **40** (2013), 110-116.
- [13] Z. Ji, F. Xu, and B. Wang, A category-integrated language model for question retrieval in community question answering, in: *AIRS*, 2012, pp. 14-25.
- [14] F. Xu, Z. Ji, and B. Wang, Dual role model for question recommendation in community question answering, in: *SIGIR*, 2012, pp. 771-780.
- [15] Z. Ji and B. Wang, Learning to rank for question routing in community question answering, in: *CIKM*, 2013, pp. 2363-2368.
- [16] Z. Ji, F. Xu, B. Wang, and B. He, Question-answer topic model for question retrieval in community question answering, in: *CIKM*, 2012, pp. 2471-2474.
- [17] J.W. Ely, J.A. Osheroff, P.N. Gorman, M.H. Ebell, M.L. Chambliss, E.A. Pifer, and P.Z. Stavri, A taxonomy of generic clinical questions: classification study, *BMJ* **321** (2000), 429-432.
- [18] R. Schnall, J.J. Cimino, L.M. Currie, and S. Bakken, Information needs of case managers caring for persons living with HIV, *JAMIA* **18** (2011), 305-308.
- [19] H. Guo, J. Li, and T. Dai, Consumer Health Information Needs and Question Classification: Analysis of Hypertension Related Questions Asked by Consumers on a Chinese Health Website, in: *MedInfo*, 2015, pp. 810-814.
- [20] J. Cohen, A Coefficient of Agreement for Nominal Scales, *Educational and psychological measurement* **20** (1960), 37-46.

Address for correspondence

Hua Xu

Hua.Xu@uth.tmc.edu

Acceptance and Use of eHealth/mHealth Applications for Self-Management Among Cancer Survivors

Yun Jiang^a, Brady T. West^b, Debra L. Barton^a, Marcelline R. Harris^a

^a School of Nursing, University of Michigan, Ann Arbor, Michigan, USA

^b Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA

Abstract

Cancer survivors' acceptance and use of eHealth/mHealth applications for self-management can be unique and are not fully understood. We used data from the Health Information National Trends Survey 4 Cycle 4 to examine cancer survivors' acceptance and use of eHealth/mHealth applications for key self-management processes, and conducted logistic regression and Rao-Scott design-adjusted Chi-square tests to assess bivariate associations between potential predictors and actual use. Potential factors were selected based on the Individual and Family Self-Management theory. High acceptance of eHealth applications was identified, and adoption of mHealth was relatively low. Younger, higher educated, married, employed, and higher income survivors tended to use eHealth/mHealth applications for self-management. Survivors who were newly diagnosed or still on treatment were more likely to look for cancer information online or communicate with health providers electronically. BMI and rural residency were associated with use of mHealth apps to achieve a health-related goal and treatment decision-making.

Keywords:

Mobile Applications; Self-Management; Neoplasms

Introduction

Advances in cancer diagnosis and treatment have led to longer survival. More than 15 million Americans diagnosed with cancer were alive by January 1, 2016 [1]. Similar to persons with other chronic conditions, cancer survivors are expected to perform self-management starting at the point of diagnosis, in order to achieve their long-term care goals. However, cancer survivors may lack the confidence or skills to perform self-management of symptoms, take medications, implement lifestyle changes, and deal with other consequences of cancer [2]. Many factors can affect their long-term engagement in self-management, such as individual, condition-specific, family, and environmental factors [3]. Conceptual descriptions of self-management have identified five core self-management processes, including problem solving, decision-making, resource utilization, partnerships with healthcare providers, and taking action [4]. To accomplish those processes, cancer survivors will need support from family and friends, healthcare professionals, communities, health systems, and possibly information and communication technologies [5].

The development of Web- and mobile-based health-related applications is increasing, with the goal of facilitating health behavior changes and support of patients in chronic disease self-management [6, 7], including limited applications in cancer self-management [8, 9]. Some applications have

integrated behavior change techniques, such as goal setting, self-monitoring, and decision support feedback, potentially making them convenient and powerful tools to support self-management behaviors and improve health outcomes [10]. However, a significant body of scientific evidence has not been established for the effectiveness of current web- and mobile-based self-management interventions [6-8, 11]. It has been suggested that the eHealth/mHealth system development process should fully assess user perceptions and address their unique needs [9]. More generally, the feasibility and acceptability of using web- or mobile-based technology for health self-monitoring needs additional investigation [7].

The purpose of this study was to understand cancer survivors' perceptions and actual use of Internet and mobile technology to support their cancer self-management needs, with a specific focus on four self-management processes: (1) resource utilization, operationalized as cancer information seeking and access to personal health information; (2) treatment decision making; (3) taking action, operationalized as achieving health-related goals; and (4) partnership with healthcare providers, operationalized as exchanging medical information with health care professionals. Further, factors associated with survivors' actual use of technology for self-management processes were explored.

Methods

Data used in the study were from the National Cancer Institute's Health Information National Trends Survey 4 (HINTS 4) Cycle 4, which is the first cycle of HINTS that included questions about adoption of mHealth applications [12]. The HINTS targets American adults aged 18 and older, to assess their knowledge of, attitudes toward, and use of cancer- and health-related information [12]. Cycle 4 used a single-mode mail survey, with a two-stage sample design, including a stratified sample of addresses and a selected adult within each sampled household [12]. The data were collected from 3,677 respondents from August to November 2014, with an overall 34.04% response rate [12]. This study focused on the subpopulation of individuals who have ever been diagnosed with cancer, of which there were 542 respondents in HINTS 4 Cycle 4.

Guided by the Individual and Family Self-Management Theory [3], potential predictive factors were categorized as contextual factors and self-management process factors, including socio-demographic characteristics, clinical characteristics (diagnosis and treatment), health status, psychological distress, knowledge and beliefs (self-efficacy, belief in cancer cause), self-regulation skills and abilities (preference of shared decision making), and social facilitation (social support and regular exercise pressure from others)

(Table 1). As shown in Table 2, outcome measures included cancer survivors' acceptance of the Internet and mHealth apps, and actual use of the Internet and mHealth apps for health self-management processes.

All statistical analyses considered the complex design of the HINTS 4 sample. The final sample weight variable was used to calculate population estimates, and 50 replicate weights were used to calculate accurate standard errors of the weighted estimates using the jackknife replication method [12]. Specifically, descriptive statistics were used to summarize the cancer survivor population characteristics and their acceptance and use of eHealth/mHealth technologies for self-management. Bivariate logistic regression analyses were used to assess relationships between each potential factor and cancer survivors' seeking cancer information via the Internet, access to online patient portals (PHR), and exchanging medical information with a health care professional through various technologies. Regression coefficients were estimated by pseudo-maximum likelihood estimation methods and odds ratios were reported. Rao-Scott design-adjusted Chi-square and *F*-tests were used for cross-tabulations to examine associations between each potential factor and cancer survivors' using mHealth apps for achieving a health-related goal, and for making a treatment decision. All statistical analyses were conducted using Stata (version 14, StataCorp LP., College Station, TX). The level of significance was 0.05.

Results

Most adult cancer survivors in the US were more than 65 years old, female, non-hispanic white, at least high school educated, currently married, unemployed, and reported their household income to be over \$50,000. More than half of survivors were overweight or obese, had the diagnosis of cancer more than 5 years ago, and had at least 2 other chronic conditions in addition to cancer. The majority felt good general health and no distress, were very confident in taking good care of their own health, and had support from their friends or family (Table 1).

Table 1 – Summary of Potential Predictive Factors (NH: Non-Hispanic, HS: High School).

Factors	% (Weighted)	Factors	% (Weighted)
Age		Race/Ethnicity	
18-34	3.8	NH- White	80.4
35-49	15.1	NH-Black	6.2
50-64	34.8	Hispanic	11.2
65+	46.3	Other	2.2
Gender		Marital Status	
Male	37.6	Married	65.2
Female	62.4	Not Married	34.8
Employment		Income	
Employed	36.6	< \$50,000	46.3
Un-employed	63.4	\$50,000+	53.7
Education		BMI	
<HS	15.7	Underweight	9.2
HS/Some College	47.9	Normal	31.1
College+	36.4	Overweight/obese	59.7
Rural		Regular Provider	
Yes	5.6	Yes	81.3
No	94.4	No	18.7
Cancer type		Co-morbidity	
Breast	15.4	0	20.7

Prostate	8.3	1	26.9
Others	76.3	2+	52.4
Years since Diagnosis		Time of Last Treatment	
< 1 Year	12.4	On Treatment	12.0
2-5 Years	19.9	< 1 Year	11.5
> 5 Years	67.8	1-5 Years	22.1
Psychodistress		5+ Years	54.3
None	71.3	General Health	
Mild	16.6	Good+	80.5
Moderate/Severe	12.1	Fair/Poor	19.5
Self-Efficacy		Social Support	
Completely/Very	67.2	A little/Not at all	9.2
Somewhat	26.3	Somewhat	15.2
A little/Not at all	6.5	A Lot	75.6
Everything Causes Cancer		Exercise Pressure	
Agree	56.6	A little/Not at all	81.6
Disagree	43.4	A lot/Some	18.4
Decision Making (moderate survival)		Decision Making (low survival)	
Self-leading	39.9	Self-leading	50.8
Shared	47.9	Shared	38.7
Doctor-leading	12.2	Doctor-leading	10.5

Regarding cancer survivors' acceptance of eHealth/mHealth applications for self-management, most survivors had access to the internet (73.9%), trusted online cancer information (68.9%), and considered it very important to be able to access to their own health information electronically (67.7%). However, more than one-third of survivors did not have a smartphone or tablet (40.4%), and only 23.4% of survivors had the mHealth apps. In addition, less than half of survivors (44.7%) reported that they were somehow or very interested in exchanging medical information with a health care professional electronically (Table 2).

In the past 12 months, a limited number of cancer survivors actually used the Internet to look for cancer information for themselves (31.8%), accessed their personal health information online (29.4%), and had exchanged medical information with a health care professional through email, text message, mobile apps, video conference or social media (29.8%). There were even fewer survivors using mHealth apps to help achieve a health-related goal (13.0%) or make a decision about how to treat an illness or condition (7.6%) (Table 2).

Table 2 – Acceptance and Use of eHealth/mHealth for Self-Management.

Measure	Level	n (% weighted)
Acceptance		
Access to the internet or send and receive e-mail	Yes	375 (73.9)
	No	154 (26.1)
Trust of online cancer information	Some/A lot	334 (68.9)
	Not at all/A little	146 (31.1)
Importance of access personal	Very	347 (67.7)

health information electronically	Somewhat	109 (19.4)
	Not at all	62 (12.9)
Having mHealth apps installed in mobile devices	Yes	92 (23.4)
	No apps	168 (36.2)
	No device	232 (40.4)
Interested in exchanging medical information with a HCP electronically	Some/Very	210 (44.7)
	Not at all/A little	260 (55.3)
Actual Use		
Seeking online cancer information in the past 12 months	Yes	147 (31.8)
	No	218 (41.7)
	No access to Internet	154 (26.5)
Access to personal health information online in the last 12 months	Yes	126 (29.4)
	No	402 (70.6)
Exchanging medical information with a HCP electronically in the past 12 months	Yes	133 (29.8)
	No	376 (70.2)
Using mHealth apps to achieve a health-related goal such as losing weight, or increasing physical activity	Yes	41 (13.0)
	No	47 (10.1)
	No apps	168 (36.4)
	No device	232 (40.5)
Using mHealth apps to make a decision about how to treat an illness or condition	Yes	31 (7.6)
	No	58 (15.6)
	No apps	168 (36.3)
	No device	232 (40.5)

As shown in Table 3, bivariate logistic regression results indicated that age, education, marital status, employment status, income, cancer type, years of diagnosis, time of last treatment, and belief that everything causes cancer significantly predicted survivors' seeking cancer information online. Specifically, the odds of seeking online cancer information were 4.02 and 2.77 times higher among survivors aged 35-49 and 50-64 years than those aged 65 years and older. The odds were also higher among higher educated, married, employed, higher income survivors. Breast cancer survivors have 2.08 times higher odds of seeking cancer information online than other cancer survivors, while the odds for prostate cancer survivors were 65% lower than other cancer survivors. Cancer survivors who were recently diagnosed or still on treatment had higher odds of seeking online information than those who had been diagnosed for more than 2 years or had their last treatment more than 1 years ago. Other factors, such as gender, race, BMI, number of comorbidities, health status, self-efficacy, and social support, were not found to be significant correlates.

Access to PHR and exchanging medical information with HCP electronically had similar associations with the socio-demographic predictors, that is, married, employed, and higher income survivors had higher odds of using eHealth applications for self-management. However, education was not found to be associated with survivors' access to PHR. Cancer survivors who were diagnosed less than 1 year ago had higher odds of using technology for exchanging medical

information with HCP. Other cancer-related characteristics, such as cancer type and years of diagnosis, and beliefs that everything causes cancer were not found to be associated with either access to PHR or exchanging medical information electronically (Table 3).

Table 3 – Significant Predictive Factors of Using eHealth Applications for Self-Management PHR (PHR: Personal Health Records, HCP: Health Care Professionals, HS: High School, OR: Odds Ratio).

Factors	Seeking Online Cancer Inform		Access to PHR		Exchanging Medical Inform. with HCP	
	F-Test		F-Test		F-Test	
	OR	p	OR	p	OR	p
Age (ref: 65+)						
	$F(3,45)=6.1$; $p<0.001$		$F(3,45)=2.5$; $p=0.07$		$F(3,45)=12.3$; $p<0.001$	
18-34	5.74	0.22	2.75	0.47	5.22	0.16
35-49	4.02	0.02	2.72	0.12	5.03	0.01
50-64	2.77	<0.01	2.31	0.01	4.20	<0.01
Education (ref: < High School)						
	$F(2,46)=3.9$; $p=0.03$		$F(2,46)=0.2$; $p=0.80$		$F(2,46)=3.5$; $p=0.04$	
HS/Some College	5.95	0.01	1.55	0.54	5.48	0.02
College+	6.71	<0.01	1.63	0.50	5.99	0.01
Marital Status (ref: Not Married)						
	$F(1,47)=12.9$; $p<0.001$		$F(1,47)=5.4$; $p=0.02$		$F(1,47)=8.0$; $p=0.007$	
Married	2.99	<0.01	2.65	0.02	2.28	<0.01
Employment (ref: Un-employed)						
	$F(1,47)=6.9$; $p=0.01$		$F(1,47)=5.9$; $p=0.02$		$F(1,47)=4.3$; $p=0.04$	
Em- ployed	2.53	0.01	2.42	0.02	2.09	0.04
Income (ref: < \$50,000)						
	$F(1,47)=21.0$; $p<0.001$		$F(1,47)=15.3$; $p<0.001$		$F(1,47)=31.2$; $p<0.001$	
\$50,000+	3.60	<0.01	4.53	<0.01	4.76	<0.01
Cancer Type (ref: Others)						
	$F(2,46)=5.9$; $p=0.005$		$F(2,46)=0.5$; $p=0.61$		$F(2,46)=0.8$; $p=0.47$	
Breast	2.08	0.05	1.54	0.32	1.62	0.25
Prostate	0.35	0.04	1.05	0.93	0.87	0.81
Years since Diagnosis (ref: < 1 Year)						
	$F(2,46)=6.01$; $p=.005$		$F(2,46)=1.37$; $p=0.27$		$F(2,46)=3.78$; $p=0.03$	
2-5 Years	0.26	0.01	0.55	0.18	0.26	0.02
> 5 Years	0.18	<0.01	0.53	0.12	0.29	0.01
Time of Last Treatment (ref: Still on Treatment)						
	$F(3,45)=7.35$; $p<0.001$		$F(3,45)=2.01$; $p=0.13$		$F(3,45)=1.23$; $p=0.31$	
< 1 Year	1.52	0.47	3.03	0.09	0.67	0.54
1-5 Years	0.24	<0.01	0.84	0.76	0.38	0.11
5+ Years	0.23	<0.01	1.03	0.95	0.41	0.09
Beliefs that Everything Causes Cancer (Ref: Agree)						
	$F(1,47)=4.66$; $p=0.04$		$F(1,47)=0.45$; $p=0.50$		$F(1,47)=0.05$; $p=0.81$	
Disagree	0.54	0.04	0.81	0.50	0.93	0.82

The assessment of using mHealth apps for self-management focused on achieving a health-related goal and treatment decision making. As shown in Table 4, younger, higher educated, married, employed, and higher income cancer survivors were more likely to use mHealth apps for self-

management. In addition, obese survivors were more likely to use mHealth apps to help achieve a health-related goal, such as losing weight or increasing physical activity, while survivors living in rural areas were more likely to use mHealth to help make treatment decisions.

Table 4 – Factors Associated with Use of mHealth Applications.

Factors	For Achieving a Health-Related Goal		For Making a Treatment Decision	
	% of YES Response	Rao-Scott F-Test	% of YES Response	Rao-Scott F-Test
Age				
18-34	62.7	$F(6.8, 319.4)=4.97$; $p<.0001$	27.8	$F(6.4, 299.3)=3.92$; $p=.0007$
35-49	30.4		20.8	
50-64	13.2		6.8	
65+	2.0		2.8	
Education				
<HS	0	$F(4.6, 215.8)=3.83$; $p=.003$	0	$F(4.3, 199.8)=4.18$; $p=.002$
HS/Some College	17.8		12.7	
College+	10.6		4.6	
Marital Status				
Married	17.1	$F(2.5, 116.9)=10.0$; $p<.0001$	9.9	$F(2.8, 133.7)=7.9$; $p=.0001$
Not Married	5.3		3.5	
Married				
Employment				
Employed	29.3	$F(2.9, 136.5)=10.8$; $p<.0001$	14.5	$F(2.7, 128.8)=10.2$; $p<.0001$
Un-employed	4.5		4.0	
Income				
< \$50,000	6.5	$F(2.8, 132.3)=15.2$; $p<.0001$	2.8	$F(2.8, 132.3)=15.8$; $p<.0001$
\$50,000+	18.6		11.6	
BMI				
Under-weight	11.5	$F(7.3, 341.9)=2.03$; $p=.048$	5.0	$F(7.2, 336.0)=1.98$; $p=.055$
Normal	10.6		4.5	
Overweight	7.0		4.5	
Obese	22.4		15.2	
Rural		$F(2.2, 104.6)=2.68$; $p=.07$		$F(2.3, 106.0)=5.14$; $p=.005$
Yes	47.8		47.8	
No	11.5		5.5	

Discussion

This study described cancer survivors' acceptance and actual use of web- and mobile-based health applications for self-management, and explored potential factors associated with their actual use. The dataset was from a national mail survey with a complex sample design. Findings of this study are expected to be generalized to the US adult cancer survivor population.

Findings of the study indicated high acceptance of eHealth applications and a relatively low adoption of mHealth apps for self-management. It is believed that the eHealth/mHealth system can provide convenient platforms for patients to engage in self-management, especially in personal health information seeking and management, health communications, and health decision support [13]. However, cancer survivors' self-management and technology adoption behaviors can be complex and have not been fully understood. This study found that less than one-third of cancer survivors used eHealth or mHealth applications for cancer information seeking, access to personal health information, or exchanging medical information with health care professionals. Cancer survivors

have a relatively lower mobile device ownership than American adults (60% vs. 68%) [14]. One possible explanation may be because that cancer survivors tend to be old (46.3% are 65+ years old). Among all cancer survivors, the percentage of mHealth apps owners is low (23%). However, among those survivors who had mobile devices, they were actually more likely to install mHealth apps than general population who have mobile devices (64% vs. 58%) [15]. The proportion of cancer survivors who used mHealth apps to help achieve health goals or treatment decision-making is similar as that in the general US adult population [16].

Although predictive factors of using eHealth/mHealth applications for self-management were slightly different depending on the type of technologies and self-management processes, survivors' technology use were mainly associated with their socio-demographic characteristics, such as age, education, marital status, employment status, and household income. Younger, higher educated, married, employed, and higher income survivors were more likely to use eHealth/mHealth applications for self-management, which is congruent with the literature [13-16]. Some cancer-specific factors were found to be associated with survivors' online cancer information seeking, and communication with health care professionals. It is understandable that newly diagnosed cancer survivors and survivors who were still on active treatment have tremendous needs for health information in order to understand the disease, and deal with consequences of diagnosis and treatment [17]. This study also found that cancer survivors who believed that everything causes cancer were more likely to look for cancer information online. This finding is inconsistent with a previous report that beliefs in everything causes cancer was not associated with cancer information seeking in general population [18]. The potential interpretation may be because cancer survivors are more motivated to obtain cancer information in order to understand their diagnosis of cancer than the general population.

Cancer survivors' use of mHealth apps for self-management was found to be associated with their BMI and rural residency. As more than half of cancer survivors were overweight or obese, and many weight loss apps were integrated in smartphone or available for free download [19], it is understandable that cancer survivors would use mhealth apps to help achieve a health-related goal, such as losing weight or increasing physical activity. Although rural residency was not found to be associated with mHealth app use in the general population [16], cancer survivors living in rural areas seem to benefit from their increased access to health services and health information through mHealth technology [20].

This study did not find any association between race and use of eHealth/mHealth applications. This is not consistent with research on the US general population, which indicates that African American have higher odds of using the mHealth application for achieving health goals and treatment decision making than whites [16]. It is unclear whether cancer health disparities play a role in survivors' use of mHealth technologies for self-management, which will need to be further explored. It is interesting to notice that self-efficacy, one of potential mechanisms indicated in many self-management interventions, was not found to be associated with cancer survivors' use of eHealth/mHealth applications for self-management in this study. This finding is inconsistent with the report from the general population [16]. Cancer survivors' self-efficacy to self-manage can vary widely according to the illness-related tasks [21]. However, cancer-related self-efficacy was not specifically measured in the HINTS survey, which was a potential limitation of the study.

There are a few other limitations in this study. First, only bivariate analyses were conducted to explore associations between potential factors and outcome variables. Some survey questions had a small number of responses from cancer survivors. For example, only 41 survivors reported that they had used mHealth apps for achieving a health-related goal. Therefore, the jackknife replication method was not able to be conducted in a multivariate regression model to calculate design-adjusted standard errors [12]. Although findings of the study tend to be preliminary, significant factors identified by bivariate analyses would contribute to the final selection of potential predictors in future multivariate analyses. Second, although a self-management theory was used to guide the consideration of potential factors for cancer survivors' use of eHealth/mHealth applications for self-management, due to the limited number of survey questions available, some constructs in the theory could not be operationalized by measures in the survey. It is also possible that some potential factors may not match exactly with theoretical concepts. Last, survey data were collected about two years ago, which may undermine survivors' technology adoption for the present day.

Conclusion

This study described cancer survivors' acceptance and actual use of eHealth/mHealth applications for self-management, and revealed high acceptance of eHealth applications and relatively low adoption of mHealth apps for self-management. Less than one-third of cancer survivors used eHealth/mHealth applications for self-management, which may be limited by availability of technologies. A few predictive factors of using eHealth/mHealth applications among cancer survivors were different from those identified for the general population, with the exception of socio-demographic predictors, such as age, marital status, employment, and income. In addition, cancer-specific factors were identified to be significantly associated with survivors' use of eHealth/mHealth applications, indicating that cancer survivors may have unique needs for their use of technologies for self-management. Future studies with a larger sample size will be needed to further explore the predictive model for cancer survivors' acceptance and use of eHealth/mHealth applications for self-management.

References

- [1] American Cancer Society, *Cancer Facts & Figures 2016*, American Cancer Society, Atlanta, GA, 2016.
- [2] R. McCorkle, E. Ercolano, M. Lazenby, D. Schulman-Green, L.S. Schilling, K. Lorig, and E. H. Wagner, Self-management: Enabling and empowering patients living with cancer as a chronic illness, *CA Cancer J Clin* 61 (2011), 50-62.
- [3] P. Ryan and K.J. Sawin, The Individual and Family Self-Management Theory: background and perspectives on context, process, and outcomes, *Nurs Outlook* 57 (2009), 217-225 e216.
- [4] K.R. Lorig and H. Holman, Self-management education: history, definition, outcomes, and mechanisms. *Ann Behav Med* 26 (2003), 1-7.
- [5] P.M. Gee, D.A. Greenwood, D.A. Paterniti, D. Ward, and L.M. Miller, The eHealth Enhanced Chronic Care Model: a theory derivation approach, *J Med Internet Res* 17 (2015), e86.
- [6] K.R. Lorig, P.L. Ritter, D.D. Laurent, and K. Plant, Internet-based chronic disease self-management: a randomized trial, *Med Care* 44 (2006), 964-971.
- [7] K.A. McMillan, A. Kirk, A. Hewitt, and S. MacRury, A Systematic and Integrated Review of Mobile-Based Technology to Promote Active Lifestyles in People With Type 2 Diabetes, *J Diabetes Sci Technol* (2016), pii: 1932296816656018.
- [8] A.R. Kim and H.A. Park, Web-based Self-management Support Interventions for Cancer Survivors: A Systematic Review and Meta-analyses, *Stud Health Technol Inform* 216 (2015), 142-147.
- [9] S. Darlow and K.Y. Wen, Development testing of mobile health interventions for cancer patient self-management: A review, *Health Informatics J* 22 (2016), 633-650.
- [10] C.H. Yang, J.P. Maher, and D.E. Conroy, Implementation of behavior change techniques in mobile applications for physical activity, *Am J Prev Med* 48 (2015), 452-455.
- [11] K. Anderson and L.M. Emmerton, Contribution of mobile health applications to self-management by consumers: review of published evidence, *Aust Health Rev* (2015), doi: 10.1071/AH15162.
- [12] National Cancer Institute, Health Information National Trends Survey, <http://hints.cancer.gov/>, Last updated 16 November 2016, Accessed 1 December 2016.
- [13] K.J. Serrano, M. Yu, W.T. Riley, V. Patel, P. Hughes, K. Marchesini, and A.A. Atienza, Willingness to Exchange Health Information via Mobile Devices: Findings From a Population-Based Survey, *Ann Fam Med* 14 (2016), 34-40.
- [14] M. Anderson, The Demographics of Device Ownership, <http://www.pewinternet.org/2015/10/29/the-demographics-of-device-ownership/>, Posted 29 October 2015, Accessed 1 December 2016.
- [15] P. Krebs and D.T. Duncan, Health App Use Among US Mobile Phone Owners: A National Survey, *JMIR Mhealth Uhealth* 3 (2015), e101.
- [16] S.S. Bhuyan, N. Lu, A. Chandak A, H. Kim, D. Wyant, J. Bhatt, S. Kedia, and C.F. Chang, Use of Mobile Health Applications for Health-Seeking Behavior Among US Adults, *J Med Sys* 40 (2016), 153.
- [17] A. Schmidt, C. Kowalski, H. Pfaff, S. Wesselmann, M. Wirtz, and N. Ernstmann, The Influence of Health Literacy on Information Needs Among Women Newly Diagnosed With Breast Cancer, With Special Reference to Employment Status, *J Health Commun* 20 (2015), 1177-1184.
- [18] L.C. Kobayashi and S.G. Smith, Cancer Fatalism, Literacy, and Cancer Information Seeking in the American Public, *Health Educ Behav* 43 (2016), 461-470.
- [19] J. Chen, J.E. Cade, and M. Allman-Farinelli, The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment, *JMIR Mhealth Uhealth* 3 (2015), e104.
- [20] H.D. Carter-Templeton and L. Wu, Using Mobile Technologies to Access Evidence-Based Resources: A Rural Health Clinic Experience, *Nurs Clin North Am* 50 (2015), 595-603.
- [21] C. Foster, M. Breckons, P. Cotterell, D. Barbosa, L. Calman, J. Corner, D. Fenlon, R. Foster, C. Grimmett, A. Richardson, and P.W. Smith, Cancer Survivors' Self-Efficacy to Self-Manage in the Year Following Primary Treatment, *J Cancer Surviv* 9 (2015), 11-19.

Address for correspondence

Yun Jiang, PhD, MS, RN

School of Nursing, University of Michigan, 400 North Ingalls, Ann Arbor, MI 48109. E-mail: jiangyu@umich.edu

Promotion of Adequate Exercise for Chronic Disorders' Elderly Through Paced Music

Gi-Ryong Kang^c, Il Kon Kim^a, Woo Jin Kim^b, Do-Youn Lee^a

^a School of Computer Science & Engineering, Kyungpook National University, Daegu, Republic of Korea,

^b Center of Self-Organizing Software-Platform, Kyungpook National University, Daegu, Republic of Korea

^c Samsung Electronics Co., Ltd. Mobile Comm., Gumi, Republic of Korea

Abstract

South Korean population has been aging at an accelerated rate in recent years and the country will be a hyper-aging nation at 2020. One important issue in an aged society is higher rates of chronic disorders among the population. One of the best means to prevent an increase in severity of chronic disease is making patients exercise regularly. In this paper, we present a mobile app for senior citizens that provides daily exercise recommendations and receive tailored service related to a chronic disorder. In this app, pace (BPM) of played music is regulated to encourage patients to fulfill their daily exercise goal. The combination was constructed by using suggestions of World Health Organization and Korea Sports Promotion Foundation. Through our app, users are guided to exercise daily and regularly with appropriate exercise intensity. Patients can deal with their chronic disorders via our mobile application with increased physical activities.

Keywords:

Elderly; Chronic disease; Physical Activity; Mobile Health

Introduction

The Organization for Economic Cooperation and Development (OECD) reported that life expectancy at birth and the amount of increase in life expectancy generally have been on the rise globally. For example, the life expectancy of a newly-born baby in 2009 is 11.9 years longer than that of a baby born in 1960 [1].

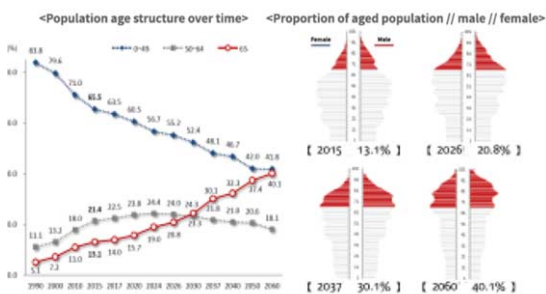


Figure 1 – South Korea's population age structure between 1990 and 2060 (Estimation)

The explosion of older population in South Korea let the rate of chronic disorders increases well as the number of years people live with health issues. Cardiovascular problems, chronic obstructive pulmonary disease, and diabetes are mostly chronic and these chronic diseases plus cancers are responsible for two-thirds of human death in the world [2].

WHO reports that people over 65 comprises 11.8% of South Korea's population and 88.5% of this population suffers from one or more diseases or disorders that are chronic [2].

Explosion of aged population let the rate of chronic disease in the population increase and would pose challenge to medical services, social welfare system, and the economy of South Korea.

This paper proposes a funny app that could help reduce chronic disorder among elderly population. It helps the user to define their exercise goals and work out with songs at measured BPM. We conducted experiments to verify if this app is effective in increasing user's exercise activity.

Methods

In 2010, World Health Organization presented Global Recommendations on Physical Activity for Health, which promotes physical activities for world's population to reduce cancer, strengthen cardiovascular system, metabolism, and musculoskeletal system.



Figure 2 – An excerpt from World Health Organization's health recommendation.

In the Recommendation, people older than 65 years of age are assumed to have limited bodily capabilities and it explicitly recommends to do as much exercise as possible although the recommended amount of exercise cannot be achieved [3].

Based on the Recommendation, South Korean government launched an organization called by 'Korea National Fitness Award 100' through Korea Sports Promotion Foundation (KSPO) and published a catalog of physical activities that help with chronic disorders [4].

Table 1 lists a few disorders that can be alleviated with exercises as recommended by WHO and KSPO. Medical research supports the selection of exercises.

Table 1 – KSPO Recommendation-based Exercise program for a few chronic disorders/diseases.

Disorder	Details
High Blood pressure	Exercise Strength: Over 60% of PssMAX, Rating of Perceived Exertion(RPE): from 11 to 13 _40 min. per session, one session per day (10 min. exercise followed by 5 min. break)

	Goal: steady increase of exercise amount based on Korea National Fitness Award 100 Exercise Recommendation for Different Disorders.
Hyperlipidemia	Exercise Strength: Above 75% of PssMAX, RPE: 12 to 16 30 min. per session, two sessions per day Goal: steady increase of exercise amount based on Korea National Fitness Award 100 Exercise Recommendation for Different Disorders.
Type 1 diabetes	Exercise Strength: Above 50% of PssMAX, RPE: 10 to 11 30 min. per session, one session per day Goal: steady increase of exercise amount based on Korea National Fitness Award 100 Exercise Recommendation for Different Disorders.
Type 2 diabetes	Exercise Strength: Above 75% of PssMAX, RPE: 12 to 16 50 min. per session, one session per day Goal: steady increase of exercise amount based on Korea National Fitness Award 100 Exercise Recommendation for Different Disorders.

In Table 1, PssMAX (Personal Speed Step MAX) is a scale to measure subjectively experienced exercise exertion, and it is calculated based on the speed, distance, time, and the number of steps measured via Samsung SmartPhone health App.

To measure PssMAX accurately and reliably, advanced medical instruments are necessary along with trained professionals.

However, this is costly, time-consuming, and hardly practical for the majority of elderly population, hence we opted for PssMAX, which can be readily estimated with a simple smartphone. Instruments that measure such indices are not only expensive but also requires trained professionals for measurement of these criteria in indices, hence RPE (rating of perceived exertion) is popularly used as an alternative.

Borg's table as shown in Figure 3 can be adopted for RPE scores. It offers the benefit of subjectively experienced exertion of exercise and it could supplement heartbeat rate-based scales. In short, RPE score can easily be used for general public to rate how strenuous a given exercise is.

The goal of the proposed app is promoting sustainable physical activities customized for those who suffer from chronic disorders. Presenting a simple fact alone that physical activities could alleviate chronic disorder does not automatically increase physical activities. This is true not only for elderly people but also most of the people in general.

This point could be addressed with elements that are entertaining. 'Entertainment' is defined as 'the action of providing with amusement or enjoyment' according to Oxford English Dictionary. Entertainment enriches life and helps sustain it meaningfully.

Entertainment is one major benefit that elderly take from information technology [5], and if an element of entertainment could be added to exercise, users are more likely to use the service. With this line of thought, we added music to help users get motivation.

The relation between the tempo of given music during aerobics exercises and change in the body of exercising individuals has been explored among elderly women living in Seoul [6].

One participant group exercised with slower songs of 66 BPM (Andante), another group with faster songs of 132 BPM (Allegro), and the control group exercised 3 times a week for 12 weeks. At the end of the period, both exercising groups had reduced body weight and body fat and increased muscle mass while the control group showed less significant body fat reduction and increased body weight.

During aerobics exercises, adequate BPM of music was found to improve oxygen uptake, respiratory exchange rate, minute ventilation, and carbon dioxide exhalation [6].

The issue is that measuring exercise capabilities requires various physiological indices. Among them, heart beat rate and maximal oxygen uptake are most widely used.

[Rating]	[Perception of effort]		
6	Very, very light	How you feel when lying in bed or sitting in a chair relaxed. Little or no effort.	
7			
8			
9			
10	Very light	Target range: How you should feel with exercise or activity.	
11	Fairly light		
12	Somewhat hard		
13			
14			
15	Hard	How you felt with the hardest work you have ever done.	
16			
17	Very hard		
18	Don't work this hard!		
19			Very, very hard
20			Maximum exertion

Figure 3 – Borg for RPE (Rating of Perceived Exertion) Scale.

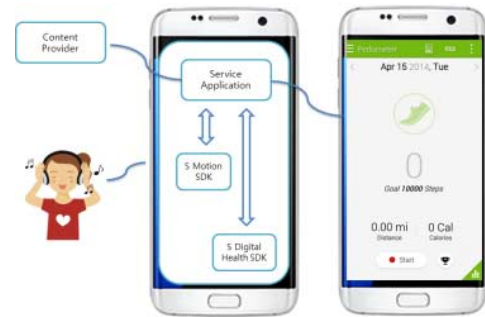


Figure 4 – Service Architecture.

The overall structure of the proposed app is depicted in Figure 4. The Application consists of the following modules: Content Provider Server, S Motion, S Digital Health, and the app itself. Service Application is the core of the service, the external Content Provider Server is responsible for providing music lists when the user is exercising, and the Application requests appropriate music list based on user's current exercise data, which is stored in the S Health DB. S Health Tracker helps browsing exercise data for hour of the day or for each date. The Content Provider offers 'oldies' songs popular among senior citizens at a low cost. Each song is tagged with BPM for different levels of exercises.

By default on the app, two music lists are created: one for songs with lower BPM and the other with high BPM. As the service gains user attraction, we believe more songs could be provided.

No known commercial music service exists that fits our description, however, as the elderly population explodes and they become more comfortable with using mobile devices, we expect to see in the future these music service apps on the market. S Motion is the solution to collect motion information from the device. It collects from the device sensors raw data related to pedometer or activity recognition or the motions associated with answering the phone.

Figure 5 shows the S Motion architecture [7], involving the following components:

- Applications: One or more applications that use Motion.
- Motion: Motion components for managing specific pedometer and activity events.
- MREngine: Motion components for providing Motion with call motion events.
- SContext: Motion components for providing Motion with pedometer and activity events.

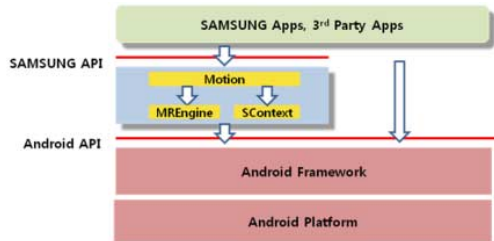


Figure 5 – Motion Architecture.

S Motion is controlled by the Application module. It takes exercise data from SensorHub module via Motion SDK Interface. The exercise data contains the number of steps taken, moved distance, time, speed, and so on. Whenever there is an update to the exercise data, the data is transmitted to the Application through Motion SDK Interface. The following modules were used in implementing our service: Smotion, SmotionPedometer, SmotionPedometer.Info Class, and SmotionPedometer.ChangeListener Interface

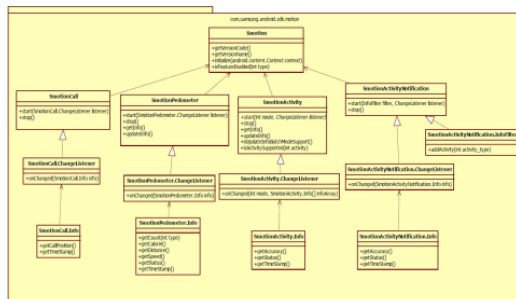


Figure 6 – Motion classes and interfaces diagram.

Samsung Digital Health helps developers to synchronize health data with S Health 4.x safely and to create useful health applications.

The health data framework keeps and provides users health data of various data types safely. Health data from a specific source device that has various sensors such as pedometer, accelerometer, or heart rate sensor is inserted based on the unified data unit, read, updated, or deleted through the health data framework.

The S Health Service package provides the tracker feature to show users health data information appropriately on S Health 4.x.[8].

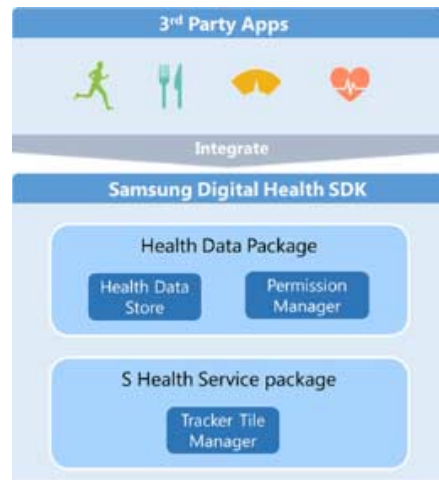


Figure 7 – Samsung Digital Health Service.

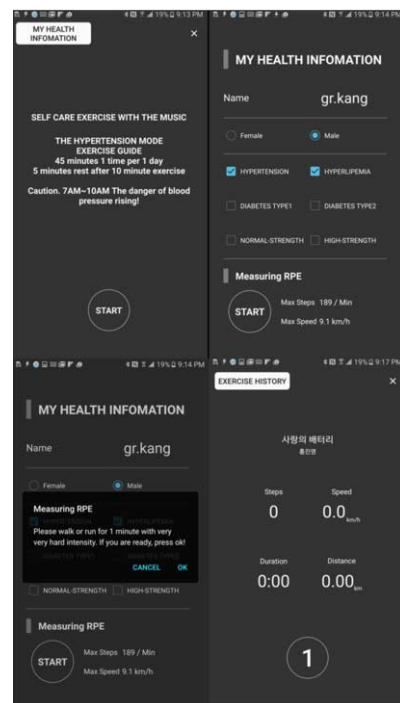


Figure 8 – App screen of Self-Care Exercise with Music.

The overall architecture of S Digital Health is shown in Figure 7. Permission to access S Health Data Type and S Health Data Store is managed with Health Data Package. The Application is granted access permission and processes the exercise data in the form of S Health Data Type when the user has paused or finished exercise and stores the data in Health Data Store database module. Our mobile app includes a service module that provides exercise data recorded through S Health Service Package. To implement this service, we used the following classes: *HealthDataStore*, *HealthPermissionManager*, *TrackerManager*, *TrackerInfo*, *HealthDataResolver*, and *HealthDevice*.

When the app is launched (Figure 8), an exercise guide is presented based on user preference as specified in the Preference menu. The user could save basic user information,

the type of chronic disease/disorder, and maximum limit for exercise exertion in 'My health data'.

Results

To test the effectiveness of our mobile app, we contacted a local church and asked elderly congregation members to participate in our experiment. A male and a female seniors aged between 65 and 70, and also a male and a female between 71 and 75 participated. Three of them had a different chronic disorder. Their profiles are shown below.

The result showed that music contributed to an increase in physical activities by the participants.



Figure 9 – PssMax data measured for participants.

Table 2 – Application usage data for participants.

Participant	RPE	Recommended pace per minute	Recommended speed	Target steps
A(Healthy)	80%	139	6.08 km/h	5560
B(High blood pressure)	80%	113	5.46 km/h	4520
C(Type 1 diabetes)	80%	96	4.3 km/h	2880
D(Type 1 diabetes)	80%	84	3.6 km/h	2520
Before using app (proportion to target steps)		After using app (proportion to target steps)		
4809 (-13.5%)		5716 (+2.8%)		
4091 (-9.5%)		5058 (+11.9%)		
3110 (+8%)		3395 (+17.9%)		
2144 (-14.9%)		2432 (-3.5%)		

As shown in Figure 9, we had our participants either run or walk fast strenuously for one minute and measured PssMAX with the our app.

Participant B had chronic high blood pressure and hyperlipidemia at the time of experiment, and we set the exercise mode to the easiest one, tailored for high blood pressure patients.

To maintain exercise exertion at 60% of PssMAX, then the person has to take 113 steps per minute at the speed of 5.46 km/h. The person has to take 4,520 steps for 40 minutes.

Participant B took 4,091 steps, which is 9.5% lower than recommended amount, without running our app, while this person took 5,058 steps, 11.9% more than recommended amount.

Table 2 shows recommended exercise goals based on each participant's PssMAX value as determined by the type of chronic disorder the participant has.

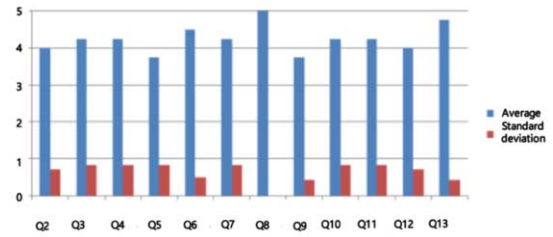


Figure 10 – Summary of PssMax survey results.

The participants were asked how satisfied they were with the service on a 1-to-5 Likert scale, where 5 means most satisfied and 1 means not satisfied at all.

Our participants responded that walking while listening to music is an effective way of exercise, and combining exercise and music makes them energetic and happy.

They also indicated that measuring maximum exercise exertion with their own smartphone rather than advanced instruments is very convenient.

Discussion

Accurate measurement of maximal exercise exertion requires costly instruments and trained professionals. In addition, the average income of elderly population is not high enough to

Table 3 – Survey items about PssMax applied experiment.

- Age:		- Gender:	
①Not likely at all ②Not likely ③Neutral ④Likely ⑤Most likely			
1	I have used an exercise management app on the smartphone.	① ② ③ ④ ⑤	
Answer the questions if you answered 'No' in Question 1			
2	This service is unique	① ② ③ ④ ⑤	
3	This service is easy to use	① ② ③ ④ ⑤	
4	Exercise management service using music is effective in this service	① ② ③ ④ ⑤	
5	Tailoring exercise for chronic disorder is effective in this service	① ② ③ ④ ⑤	
Answer the following questions if you answered 'Yes' in Question 1			
2	This service is different from other exercise management services	① ② ③ ④ ⑤	
3	This service is easier to user than other services	① ② ③ ④ ⑤	
4	Exercise management service using music is more effective in this service than other similar services	① ② ③ ④ ⑤	
5	Tailoring exercise for chronic disorder is effective in this service in comparison with other services	① ② ③ ④ ⑤	
6	Walking while listening to music with this service is enjoyable	① ② ③ ④ ⑤	
7	Walking as an exercise becomes more effective while listening to music with this service	① ② ③ ④ ⑤	
8	Setting the maximum exercise exertion level with this service is convenient	① ② ③ ④ ⑤	
9	Maximum exercise exertion level seems accurately measured	① ② ③ ④ ⑤	
10	This service can help me achieving daily exercise goal and forming a daily habit of exercising	① ② ③ ④ ⑤	
11	I will sign up for this service when it is officially launched	① ② ③ ④ ⑤	
12	I will use this service every day after it is officially launched	① ② ③ ④ ⑤	
13	I will recommend this service to others.	① ② ③ ④ ⑤	
14	Your comment:		

justify the cost, hence such devices are not easily accessible in everyday use for them. PssMAX offers only a ballpark figure, which is its most obvious limitation.

Just with a smartphone, numbers such as heartbeat rate, blood sugar level, and blood pressure cannot be easily measured. However, many tools that measure these metrics are on the market as wearable devices. Once such devices are connected with smartphones, health apps such as ours could utilize them for improved measurements. Speed, distance, and time can be obtained with the sensors built in smartphones. Certain metrics can be measured more accurately in the future with further development of smartphones.

It is true that short time heavy exercise is more effective than continuous light load exercise, and it is also included in ACSM's physical activity recommendation (over 65 years old) [9].

However, it is important that ACSM recommends increasing the maximum amount of exercise load and time that elderly can do if they are unable to perform heavy exercise because of chronic disease and physical limitations.

Conclusion

The main contribution of our research is that users are guided to exercise with sufficient exertion by using an easy-to-measure scale of PssMAX. Eventually, users' chronic disorders can be managed properly through increased physical activities.

Our mobile app is different from other existing health apps by providing exercise plans that reflect the user's exercise profile and chronic disease/disorder.

In the future work, more sensors could be added to collect user's exercise data. Music selection could be further tailored for different types of chronic disease/disorder.

Acknowledgements

This work was supported by the Technology Innovation Program (10053584, Standardization of Human Genome Sequencing Report in Electronic Medical Record System) funded By the Ministry of Trade, Industry & Energy (MI, Korea) and the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005)

References

- [1] Ministry of Health and Welfare, OECD Health Data 2015 (2015), 4-6.
- [2] World Health Organization, Increasing Levels of Physical Activity (2010), 10.
- [3] World Health Organization, Increasing Levels of Physical Activity (2010), 15.
- [4] Korea Sports Promotion Foundation, Korea National Fitness Award 100- Exercise Program for Purpose (2016),
- [5] J.U. Ko and J.O. Chang, A Research View of Developmental Task on the Entertainment Industry through the Internet Use of the Old (2011), 299.
- [6] B.G. Choi and H.K. Yoon, The Influence of Music Tempo on Body Composition and Functional Fitness for Elderly Women during Aerobic Exercise, *Korean journal of sports science* **21** (2) (2012), 1061-1062.
- [7] Samsung, Motion Programming Guide, <http://developer.samsung.com/galaxy/motion> (Mar 25, 2016).
- [8] Samsung, S Health Branding Guidelines with Samsung Digital Health SDK, <http://developer.samsung.com/health> (Sep 7, 2016).
- [8] Samsung, S Health Branding Guidelines with Samsung Digital Health SDK, <http://developer.samsung.com/health> (Sep 7, 2016).
- [9] ACSM's Exercise is Medicine (2009), 10.

Address for correspondence

Corresponding author:

Il Kon Kim

Computer Science and Engineering Graduate School, Kyungpook National University, Daegu, Republic of Korea

E-mail: ikkim@knu.ac.kr

Author:

Gi-Ryong Kang

Computer Science and Engineering Graduate School, Kyungpook National University, Daegu, Republic of Korea

E-mail: kgr9601@daum.net

Woo-Jin Kim

Center of Self-Organizing Software-Platform, Kyungpook National University, Daegu, Republic of Korea

Daegu, Korea Republic of,

E-mail: yunaloving@gmail.com

Do-Youn Lee

Computer Science and Engineering Graduate School, Kyungpook National University, Daegu, Republic of Korea

E-mail: keveni@naver.com

Development of Parkinson Patient Generated Data Collection Platform Using FHIR and IoT Devices

Dae-Young Kim, Sun-ho Hwang, Min-Gyu Kim, Joon-Hyun Song, Sin-Woong Lee, Il Kon Kim

School of Computer Science & Engineering, Kyungpook National University, Daegu, Korea

Abstract

Internet of Things (IoT) devices can be effectively used in cases where continuous observation of patients is required, such as Parkinson's disease. This is due to the characteristics of the IoT (Internet of Things), which allows data to be measured and transmitted at any time, anywhere. In this study, we developed a health data collection platform that stores and transmits the foot pressure data of Parkinson patients using FHIR (Fast Healthcare Interoperability Resources). The platform can be used to collect the foot pressure of a large number of Parkinson's patients. Based on the accumulated data, it is possible to identify abnormal behaviors such as walking pattern, tilt and instability of stride length in patients with Parkinson's disease. Based on the results of this study, accurate diagnosis and treatment of Parkinson's disease can be made.

Keywords:

Electronic Health Records; Parkinson Disease

Introduction

IoT device can measure various data regardless of time and places. Thus individual who uses IoT devices can easily check and store data. At the same time, the postwar generation starts to retire and preventive medicine has advanced a lot recently. This results in increase of interest in health and desire to manage one's own health data.

Thanks to this situations IoT healthcare device market is growing at rapid pace. IoT healthcare device installation expected to increase greatly from 46 millions in 2015 to 161 millions in 2020 [1]. With this, absolute amount of data measured by IoT healthcare device is growing fast

This is consistent with the current situation in which big data and machine learning are in the spotlight. However, the absolute amount of health data has increased, but there are very few data that can actually be used for big data and machine learning.

The main reason for this is the absence of a standard health data collection platform. Some manufacturers collect health data which is measured in their products through their own cloud services, but their system rarely follow international standards.

Machine learning and big data are very sensitive to data types and formats. If health data collection platform collects data without using a international standard, data from the platform can not be used for machine learning and big data because most of user of the platform enters data without conforming to the data type and format specified by the platform.

Health data collection platform with FHIR can solve these problems above. It's easier to inform platform users with FHIR because it's international health data standard. Validation of data is possible with FHIR, so it is possible to sort out data that does not conform to the standard and obtain good quality data. As a result, this guarantees health data interoperability.

In this study, we would like to develop a standard platform that collects and stores health data measured by IoT healthcare device and provide anonymized data by RESTful API method. To this end, we adopted the FHIR, which is being developed as a next-generation medical information standard in HL7, as a data transmission standard. We anonymized health data because it contains sensitive personal information. We hope that this anonymized health data can be used in various studies. In order to actually apply this platform, we collected health data measured by the foot pressure sensor.

The foot pressure sensor is used to identify patterns of behavior such as abnormal stride and tilt in Parkinson's patients. The measured data can be used to compare the patient's foot pressure data with the patient's condition. A steady state analysis of Parkinson's patients and ongoing feedback are helpful in improving the patient's abnormal behavior pattern [2].

In this study, we also present how the standard platform can contribute to continuous observation and feedback of Parkinson's patients. Based on this, we would like to show how a health data platform can provide services to data providers.

Methods

Related Studies

HL7 has been developing FHIR (Fast Healthcare Interoperability Resources) as a next-generation medical standard. FHIR is suitable for IoT devices and wearable healthcare devices with low processing power due to less usage of data used for health data transmission than other medical standards such as CDA [3].

LOINC is common language for clinical and laboratory observation. It's aim is to provide universal codes and names that provide the global lingua franca for identifying tests and observations [4].

HAPI FHIR is an open source FHIR library led by the University Health Network. This is a Java-based library, so it works well with the Java Spring framework.

When disclosing data, personal identification information such as social security number or name must be removed to protect

privacy. The data anonymization method can be classified into simple anonymization through elimination of semi-identifiers, heuristic anonymization in which detailed information is hidden through some rule or human judgment of the values corresponding to semi-identifiers, anonymization through privacy model and algorithm [5].

Foot pressure sensor : Parkinson's disease is accompanied by several abnormal behaviors (slow motion, faceless face, etc.) due to the loss of neurons in the brain [6]. In addition to biopsy, behavioral analysis can briefly confirm the patient's condition. We developed and applied a foot pressure sensor to monitor the abnormal behavior pattern for a long time and to identify the gait that can form more accurate diagnosis and prescription through big data formation for the patient.

Health data collection platform

IoT healthcare devices can measure various types of health data. Examples of data that can be measured include Pulse and Oxygen in Blood (SpO2 and Heart rate), Electrocardiogram (ECG), Blood pressure, Electromyography (EMG), Glucometer, Body temperature etc. In this study, the foot pressure was measured and applied to a standard platform.

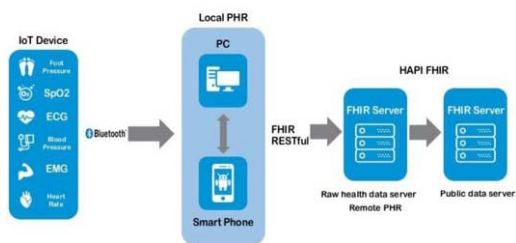


Figure 1- platform overview

An overview of the standard health data collection platform is shown in picture above. The platform consists of IoT Device, Local PHR, Raw health data Server (Remote PHR), and Public data Server.

IoT devices are connected to a PC or smartphone that acts as a local PHR via Bluetooth to transmit measured health data. The user can send the data stored in the local PHR to his raw health data server with his consent. Local PHR generates user health data in FHIR standard format and transmits data to Raw health data server. When converting health data into FHIR format, LOINC code was used as a standard for measured values and measurement positions.

Raw health data server acts as a remote PHR repository. The user can store and query health data only for his / her data, and the server provides functions for this to the RESTful API. Then the raw health data stored by the user is transferred to the public data server through the anonymization process.

In this study, data is anonymized by simple anonymization by removing the quasi-identifier. We generated a random string of 14 digits consisting of integers from 1 to 9 and letters a to z and used it as an identifier for the anonymous user (ISO/TS 25237:2008 Health informatics - pseudonymization). The original identifier and the identifier that is being anonymized are managed as a separate table in the database of the raw health data server. This allows the measurement data for the same user to be continuously updated.

Then, public data server converts the health data stored in an anonymized form into the FHIR format and then releases it to the public through the RESTful API.

Both Raw health data server and public data server were developed as Java Spring framework on Heroku cloud service and HAPI FHIR Library is used for manipulating FHIR resources.

Foot pressure sensor

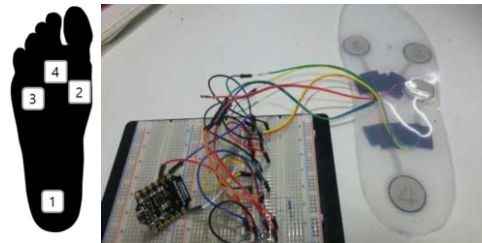


Figure 2 - foot pressure sensor structure

The foot pressure sensor is attached to the sole of the shoe by four pressure sensors, and the user's foot pressure is self-measured. The position of each sensor is shown in Figure 2 [7].

The pressure data measured at the foot pressure sensor is stored in the Android application, which is the patient's local PHR repository. The stored data is converted to the FHIR format when the user agrees, and then transmitted to the Raw health data server.

Foot pressure data of 10 healthy persons and virtual foot pressure data of 10 patients with Parkinson's diseases were measured using a foot pressure sensor. The gait of patients with virtual Parkinson's diseases was measured by imitating the gait by referring to the article [8] that studied the characteristics of the gait of the patient with Parkinson's disease. Each individual walked 10 seconds on a plain without slope and transmitted foot pressure data to the Raw health data server at 0.1 second intervals.

In order to store the measurement data in the Raw health data server, it is necessary to store the measurement data after registering the patient's basic information. Patient information uses Patient resource of FHIR and measurement data uses Observation Resource. The Raw health data server then anonymizes the data and sends it to the public data server.

Parkinson analysis from foot pressure data provided as public data

The public data server uses the Bundle Resource of the FHIR to provide public data in chronological order. Researchers studying Parkinson's disease can use anonymized user information and foot pressure data from a public data server in a variety of ways.

In this study, we derived data on the order of movement of the center of gravity and the distribution of plantar pressure from the foot pressure of anonymous Parkinson's patients and healthy persons. The center of gravity according to the movement was compared with the time when the first pressure was applied to the four sensors, and the pressure distribution of the sole was compared with the intensity of the pressure applied to each of the four sensors. As a result, we could find the difference between the foot pressure change and the foot pressure distribution in general and Parkinson's patients.

Provide services to the patient based on the analyzed data

To provide an example of the use of health data collection platform, we have developed an application that provides a service to compare the data of Parkinson's patients and the general public based on the above analysis and to analyze the progress of the treatment.

The application compares and analyzes the foot pressure data from the IoT healthcare device of Parkinson's patients against the general foot pressure data provided by the public data server. On the basis of this, the medical staff can know the current condition through the self-measurement of the patient, and can simply summarize the opinion on the progress of the treatment compared with the past patient's record. In case of an emergency, user can contact the doctor in charge. The information on the treatment progress and the bio-signal is stored in the local PHR repository so that the physician and the guardian can read it at any time.

Results

The FHIR Patient resource is generated to transfer patient information from the local PHR to the Raw health data server is as follows.

```
{
  "resourceType": "Patient",
  "identifier": [{
    "value": "910523-2895112"
  }],
  "name": [{
    "use": "official",
    "family": "Anderson"
  }],
  "given": [
    "Rollingstone",
    "K" ] },
  "telecom": [{
    "system": "phone",
    "value": "01057524885",
    "use": "work" }],
  "gender": "male",
  "birthDate": "1991-05-23"
}
```

Figure 3 - Patient resource

The resident registration number was used as an identifier of the patient and basic information such as name, phone number, sex, date of birth, etc. was stored.

Observation resource is generated to send the foot pressure measurement data from the local PHR to the raw health data server.

The patient identification number was used to identify the measurement value of the patient and the data was stored in chronological order including the measurement time. The Loinc code for the Body site and Component is not mapped yet.

Public data server uses FHIR Bundle resource to provide the anonymized foot pressure data measured at the foot pressure sensor. Due to the characteristics of the Bundle resource, multiple components can be stored, allowing the researcher to transfer large amounts of health data at a time.

```
{
  "resourceType": "Observation",
  "id": "foot-pressure",
  "identifier": [{
    "value": "910523-2895112" }],
  "effectiveDateTime": "2016-10-30T11:23:21+06:00",
  "bodySite": {
    "coding": [{
      "system": "",
      "code": "",
      "display": "Right foot sole" } ]},
  "component": [{
    "code": {
      "coding": [{
        "system": "",
        "code": "",
        "display": "foot pressure sensor 1" } ]},
    "valueQuantity": {
      "value": 116,
      "unit": "lb" } } ]}
}
```

Figure 4 - Observation resource

FHIR StructureDefinition resource describes a structure - a set of data element definitions, and their associated rules of usage. This resource is used to describe the underlying resources, data types defined in FHIR, and also for describing extensions, and constraints on resources and data types.

The base FHIR specification describes a set of base resources, frameworks and APIs that are used in many different contexts in healthcare. However there is wide variability between jurisdictions and across the healthcare ecosystem around

```
<?xml version="1.0" encoding="utf-8"?>
<StructureDefinition xmlns="http://hl7.org/fhir">
...
<base
value="http://hl7.org/fhir/StructureDefinition/Patient" />
<differential>
  <element>
    <path value="Patient" />
    <type>
      <code value="Patient" />
    </type>
  </element>
  <element>
    <path value="Patient.identifier" />
    <definition value="An identifier for this patient" />
  </element>
  <element>
    <path value="Patient.identifier.id" />
    <representation value="xmlAttr" />
    <definition value="unique id for the element within a resource (for internal references)" />
  </element>
...
</differential>
</StructureDefinition>
```

Figure 5 - Patient StructureDefinition

```

<?xml version="1.0" encoding="utf-8"?>
<StructureDefinition xmlns="http://hl7.org/fhir">
...
<base
value="http://hl7.org/fhir/StructureDefinition/Observation"
/>
<differential>
  <element>
    <path value="Observation" />
    <type>
      <code value="Observation" />
    </type>
  </element>
  <element>
    <path value="Observation.identifier" />
    <definition value="A unique identifier for the simple
observation instance" />
  </element>
  <element>
    <path value="Observation.identifier.value" />
    <definition value="The portion of the identifier typi-
cally displayed to the user and which is unique within the
context of the system" />
    <exampleString value="123456" />
  </element>
...
</differential>
</StructureDefinition>

```

Figure 6 - Observation StructureDefinition

practices, requirements, regulations, education and what actions are feasible and beneficial. For this reason, further adaptation is required to particular contexts of use.

As a consequence, StructureDefinition resources for Patient and Observation resources are defined are outlined in Figures 5 and 6.

These two StructureDefinition resources not only represent a structure constrained or extended from basic FHIR specification, but also can be used to validate FHIR resources being ex- changed for the purpose of enhancing the degree of interoperability.

Data were provided as shown in Figure 6 and Figure 7 with public health data on foot pressure. We record the point at which the pressure value applied to the foot pressure sensor initially has a value at the initial value (0) to analyze the pattern of the gait pattern. In general, the pattern of walking changes regularly from 1 → 2 → 3 → 4, but the patterns of Parkinson's patients are not known because Parkinson's pressure is applied in an irregular order [9]. We found differences in gait patterns among the general population and Parkinson's patients. Also, by comparing the pressure of each sensor at the same time, the difference in foot pressure distribution between the general and Parkinson's patients was also found.

Figure 7 is a graph of the mean of the pressure values applied to the four sensors over time while a person is walking. In the case of a normal person, we can see that the movement of weight is sequentially performed from sensor No. 1 to sensor No. 4.

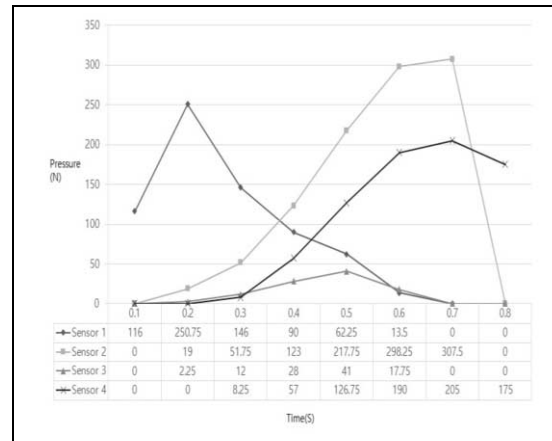


Figure 7 - Steps of healthy person

Figure 8 is a graphical representation of the mean value of pressure applied to four sensors over time while a Parkinson patient is walking. In the case of Parkinson's patients, pressure is uniformly applied to the remaining sensors except the second sensor. This can be interpreted as an unusual step where the center of gravity is constantly biased forward. Using the difference in gait patterns of these patients, the current state of the patient can be diagnosed by comparing with the normal control group.

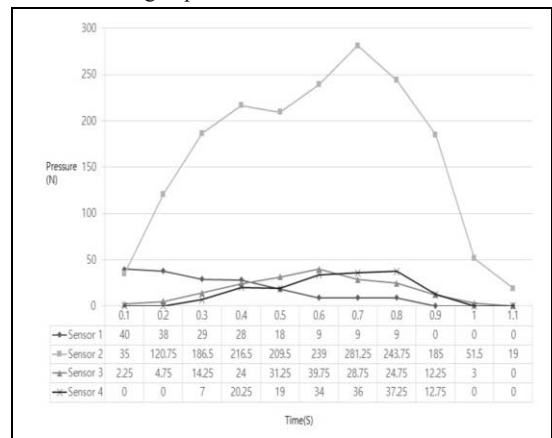


Figure 8 - Steps of Parkinson's patient

Discussion

Using the health data collection platform developed in this study, a large amount of accumulated health data can be utilized by researchers and medical personnel in various researches. Our ultimate goal is to provide beneficial services based on the results of research using this platform, and to contribute to the promotion of human health.

From this point of view, applying this platform to actual Parkinson's patients is expected to provide customized services to patients with Parkinson's disease based on the results of using the foot pressure data. The mobile application can help patients to quickly identify their condition, instantly adjust the supply of drugs under test, and help them improve treatment methods, such as understanding the course of treatment based on objective health data.

In order to implement this, there are additional things to be to the platform : linkage between hospital system and raw health data server, advanced data anonymization algorithm.

Improvements in terms of IoT devices include measuring heart rate, blood oxygen saturation, EMG, blood pressure, and various other health data to enable more comprehensive patient status confirmation.

As these improvements are implemented and the actual data of Parkinson patients are constantly accumulated on the platform and enough data is formed, it is possible to make more advanced research and diagnosis. Based on this, it will be possible to perform gait correction and early diagnosis of Parkinson's disease and correct treatment can be done.

Conclusion

The reason why many health data are not used in the medical field is that the health data is collected without using the international standard. If the data is collected without conforming to the international standard, it is impossible to verify the input data, and the collected data shows increased errors. In addition, all does not share standards set up individually, so additional processing of the data is required when the data is utilized because of the lack of versatility.

To solve this problem, this study proposed health data collection platform that collects and provides health data using FHIR, the next generation medical standard. In order to apply the platform to the medical field, we dealt with patients with Parkinson's diseases and measured the patient's footsteps through the foot pressure sensor.

Acknowledgements

This study was carried out as a result of the research project of Ministry of Science, ICT and Future Planning, Institute for Information & communications Technology Promotion and SW centered university (R2215-16-1004)

This work was supported by the Future Flagship Program(10053249, Development of Personalized Healthcare System Exploiting User Life-log and Open Government Data for Business Service Model Proof on Whole Life Cycle Care) funded by the Ministry of Trade, Industry & Energy(MOTIE, Korea).

References

- [1] Tratica, *Healthcare Wearable Device Shipments to Reach Million Units Annually by 2011*, April 14, 2016
- [2] E.J. Protas, Gait and step training to reduce falls in Parkinson's disease, *NeuroRehabilitation* (2005), 183-190.
- [3] B.Franz, A. Schuler, and O. Krauss, Applying FHIR in an Integrated Health Monitoring System, *EJBI* **11** (2015), 51-52.
- [4] About LOINC, Available at: <http://loinc.org/background>
- [5] Y. Kwon, J. Yeon, and S. Lee, Anonymization Techniques Suitable for Real Medical Datasets, *Proceedings of the Korean Information Science Society Conference Volume 38* (2011), 80-83.
- [6] Division of General Internal Medicine, University of California, San Francisco (UCSF), School of Medicine 94143-0320, USA, Parkinson's disease diagnosis and treatment, *Western journal of medicine* (1996), 234-240.
- [7] S.J. Morris, Shoe-Integrated Sensor System for Wireless Gait Analysis and Real-Time Feedback, *Proceedings of the Second Joint EMBS/BMES Conference*, (2002).
- [8] J. Chae and H. Cho, The Research of Gait on Parkinson's Disease, *Journal of the Korean Society of Physical Medicine* **4**(4) (2009), 249-255.

- [9] S. L. Joon and S.L.D-K. Shin, Features Extraction for Classifying Parkinson's Disease Based on Gait Analy, *Korean Society for internet information* **11**(6) (2010), (13-20).

Address for correspondence

Il Kon Kim, Professor, Computer Science and Engineering Graduate School, Kyungpook National University, Daegu, Korea Republic of, E-mail: ikkim@knu.ac.kr

Developing a Mobile Wellness Management System for Healthy Lifestyle by Analyzing Daily Living Activities

Chae-Gyun Lim, Zae Myung Kim, Ho-Jin Choi

School of Computing, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea

Abstract

It is important to protect health and improve quality of life for people, without causing them inconvenience in today's world. Since most people are living a busy life dealing with various activities at work, school, or home, there is a need for systematic analysis of their life patterns. However, since person's life patterns could change depending on ambient environmental factors, an effective management scheme to specify one's state is required. We propose a method, in this paper, to support and enhance the personal healthy life patterns by analyzing the daily life data that has been continuously recorded by wearable sensors, such as activity trackers. We implement a mobile wellness management system by learning RNN-based user's lifestyle model, and developing behavior recommendation using greedy policy. We also consider user context and feedback to personalize each user's lifestyle.

Keywords:

Health promotion; Life Style: Fitness Tracker

Introduction

Recently, Quality of life (QoL) is an important concept for general public, as well as researchers. Researchers have increasingly conducted research to improve QoL and daily lives of people. The QoL is defined as a subjective assessment of individual's well-being that includes both physical and psychological aspects of daily living [13]. Specifically, in case of patients with chronic diseases, such as diabetes, QoL in the short-term could be poor and may affect the motivation to manage one's own health in the long-term [10, 15]. In such situation, it is required to provide appropriate treatment concerned with not only long-term, but also short-term aspects.

As quality of daily living is related to QoL, it is frequently assessed in various questionnaires of psychometric assessment tools [6, 13]. However, most people have different quality of daily living due to the diversity of their living environment. To analyze and model each individual's lifestyle, we need to monitor entire sequences of activities of the person's daily life over time. Nowadays, there are many wearable sensors to collect behavioral and physiological data from users' daily living activities. Some researches have used these sensors that collect physiological signals, such as blood pressure, heart rate, heart rate variability (HRV), and skin conductance, in order to recognize user's affects [2; 5; 7; 12; 14].

Activity trackers, like pedometer, are useful tools to support tracking movements and manage users' activities by utilizing a cloud service on Web or mobile app. In general, activity trackers are wearable devices, such as wristband or belt, and they send various types of collected data to a smartphone and/or a server [6]. The stored data can be visualized in various ways, such as statistics of movements or graphical charts showing trends, to motivate users to reach their goals.

Today, activity trackers track multiple types of data, including

step counts, movement speed, distance, heart rate, and sleep patterns. Sometimes, exercise or dietary habits can also be recorded on a mobile app or Web site that manufacturers support. Fitbit is one of the famous brands in activity tracker available in the market. It offers many kinds of wristband-type devices that can count steps, floors and distance that user has moved, and also measure heart rate and active time.

Other famous products are Jawbone devices. Figure 1 shows one such wristband activity tracker, Jawbone UP3. It has electrodes to sense user's bio signals and a band surface to indicate small icons to display various states, such as active state, sleep mode, or notification arrival. Jawbone device and mobile app work on several smartphone platforms, and provide attractive features to track user's movements, heart rate, sleep patterns, and personal records on diet and activity. It can also indicate detailed information including awake time and the level of sleep (deep or light sleep).

Therefore, we choose the Jawbone's activity trackers for our experiments. Two main reasons of our choice are as follows:

1. They apply three bio-impedance signals, such as heart rate, respiration, and galvanic skin response (GSR), to collect data of user's activities; and
2. They track not only daily activities, but also detailed levels of patterns while the user is sleeping.

In order to utilize more useful services in combination with existing healthcare services, there have been various attempts to model the state of physical information of a particular user along with underlying measured values from various sensors and devices. Pande et al. [9] and Bouarfa et al. [3] conducted a study to model the energy consumption of a person according to the activity, and Cheng et al. [4] performed a study to model changes in heart rate during the course of various activities. Austin et al. [1] and Lipton et al. [8] attempted to accurately model and predict user's heart rate; thereby, identifying medical important situations, such as heart failure.



Figure 1– Jawbone UP3 activity tracker

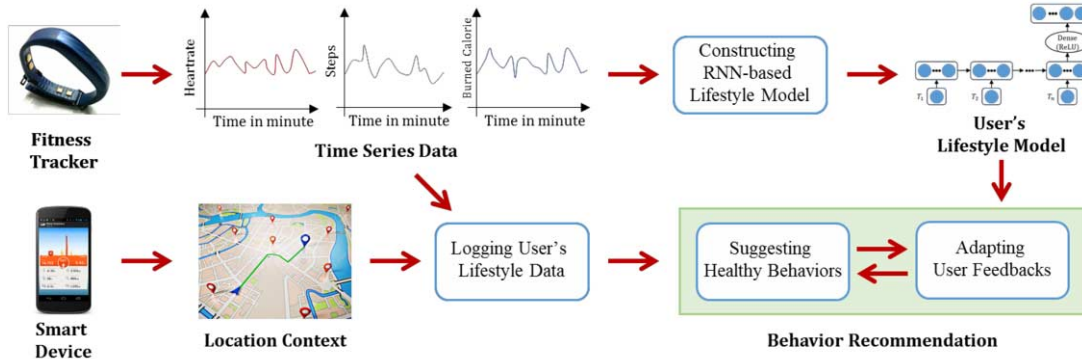


Figure 2 – Conceptual diagram of mobile wellness management system

However, this research has focused on analyzing the user state, on the basis of the present time, based on data already collected from an external device, such as a wristband-type sensor. It is, however, difficult to predict a future activity or change in situation for the user in order to guide a systematic health promotion plan. If user's activity pattern is predicted through user modeling using machine learning technology, such as DNN, it is possible to suggest proper activities at an appropriate time. Also, in order to automatically determine whether the activity is performed without causing inconvenience to the user, it is necessary to be able to observe the change of the user's activity or situation at the relevant time, and obtain feedback on the results of the activity recommendations.

In this paper, we propose a personalized wellness management system based on mobile environment to improve lifestyle quality by analyzing the recorded time series data in daily life. Figure 2 shows overall concept of our proposed system. First, we collect three types of time series data, such as heart rate, step counts, and amount of burned calories. We also log the changes of the user's location context. Based on these collected data items, we analyze and extract user's lifestyle patterns that will be used as input features. To construct an appropriate model to represent the characteristics of user's daily lifestyle, we apply a recurrent neural network (RNN) model. Lastly, we recommend healthy behaviors for the specific user, underlying the lifestyle model built in previous step. For each recommendation, the user can give particular feedback to adapt the RNN model for supporting personalization to the user's lifestyle.

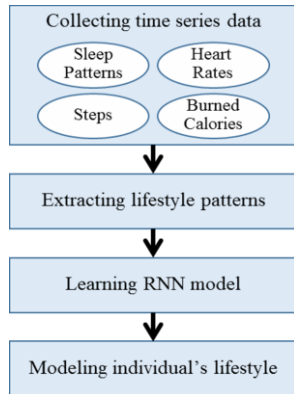


Figure 3– Block diagram for modeling lifestyle patterns of a particular user

Methods

In this section, we explain overall architecture of our mobile wellness management system in detail (Figure 3). First, we describe how our system analyzes user lifestyle, and models an individual's lifestyle patterns based on the time series data that have been obtained by wearable sensors. Second, we explain a process of activity recommendation to enhance user's wellness by utilizing the lifestyle model. Third, we introduce functionality of our mobile system, and illustrate user perspective using Android platform.

Collecting time series data

The wearable device, Jawbone UP3, collects and provides four types of time series data (i.e. sleep pattern, heart rate, step count, and amount of burned calories) from the user who is attached to the device. While only sleep patterns have nominal values over time, such as awake, light, or deep sleep, other three series have numeric values. In the data scheme of UP3, sleep pattern and heart rate are managed by different entities separately; however, both step count and burned calories are stored in a single entity named as 'moves'. These are written in JSON (JavaScript Object Notation) format. Here we show partial examples for each data gathered by UP3.

- Sleep pattern:

```
- [{"depth":1, "time":1448227726}, {"depth":2,
"time":1448228457},
...,
{"depth":2, "time":1448244340}, {"depth":3,
"time":1448245505}]
```

- Heart rate:

```
- [{"time_updated":1460041251, "resting_heartrate":62,
"sleep_ranges":[[1459960311, 1459993298]],
"bg_move_day_hr_ticks":[{"hr":67, "time":1459955996},
{"hr":69, "time":1459993357},
...,
{"hr":55, "time":1459993228}],
"xid":"piqj7ljQzD8o4kfUacmm1NTWakPuB x_",
"type":"heartrate"}]
```

- Step count and burned calories:

```
- [{"time_completed":1445712639, "distance":3,
"calories":1.15767812729, "steps":5, "time":1445712576,
"speed":0.047619048506,
...,
{"distance":2, "time_completed":1445782637,
"calories":1.17605400085, "steps":7, "time":1445782573,
"speed":0.03125}]
```

Extracting lifestyle patterns

The raw data from UP3 should be preprocessed before fitting the RNN model. Due to the difference in data type, we analyze three time series data, i.e., heart rate, step count, and burned calories. The sleep patterns are used for context information. We create three separate vectors for each collected data. The changes of user's location context provide potential clues regarding daily activity and lifestyle pattern of the user. For example, an officer's activity logs may show that there is a regular pattern between working place and home, such as "this user arrived to the working place at 9 a.m. everyday."

The patterns of user's daily lifestyle can be extracted as a multi-dimensional vector consisting of statistical metrics (sum, average, min, or max for every minute) of the collected data. Because our goal is to characterize a daily pattern for the user, we limit the data size to 24 hours while calculating features. In addition, we thought that durations shorter than one minute (e.g., every second) are not useful to represent a daily activity. These data vector patterns are used as input to the RNN model.

Learning RNN model

The RNN model is an artificial neural network, which has a directed cycle between units [11]. This model is appropriate due to the characteristics, where each unit of RNN has a time-varying activation with real-value. Figure 4 shows the RNN structure used to build the lifestyle model. In our model, each hidden layer will be calculated from each value of the given time series data T , where the number of time steps is n and i -th time step is T_i . The model used the dense function with ReLU (Rectified Linear Unit) when obtaining the output values in the last layer. By learning the three individual feature vectors (i.e., heart rate, step count, and burned calories), we can construct personalized lifestyle model for the particular user's daily living activities.

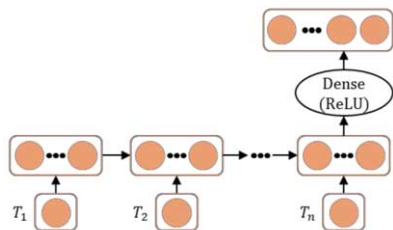


Figure 4 – RNN model to learn each time series data

Modeling individual's lifestyle

We obtain quality of lifestyle for the user by combining the collected activities of daily living and the outputs of learned RNN model. This RNN-based lifestyle model is personalized to the user's lifestyle. Therefore, each lifestyle model can be a baseline of the user's wellness, and can be utilized to make recommendations in the next stage.

Behavior recommendation

Next, our system have to decide adequate healthy behaviors based on the user's lifestyle model. Because the user's life depends on his or her own lifestyle patterns, considering context information of the user, such as location or time availability, facilitates to suggest proper healthy behavior to the user.

Figure 5 shows a flow chart illustrating procedures of behavior recommendation and adaptation to the user. This flow starts after constructing RNN-based lifestyle model for the user. First, we can obtain the user's upcoming activeness based on the lifestyle model. For instance, three types of values — heart rate, step count, and burned calories — can be predicted for the upcoming one hour from the current moment. We, then, check whether the predicted values of activeness are lower than the thresholds

corresponding to each data type. If it is true, our system generates new behavior recommendation for the user to increase the amount of future activity.

Before suggesting new behavior, we consider the user's current location context to choose what activities are suitable. For example, walking around is sufficient if the user is just coming from the lunch near the workplace, and he or she must return to work soon. Moreover, we apply greedy policy for individuals to make a decision that provide healthy behaviors. The user can send feedback to our system corresponding to the behavior recommendations in the past. There are two types of feedback: one is a flag of satisfaction, and another is whether recommended behavior is finished or not. If the user responds that he or she is not satisfied with that particular recommendation, then parameters of greedy policy in the user's context are decreased. It means that the possibility of the behavior recommendation in the specific context is also reduced, and vice-versa. This adaptation process supports better personalization for the user.

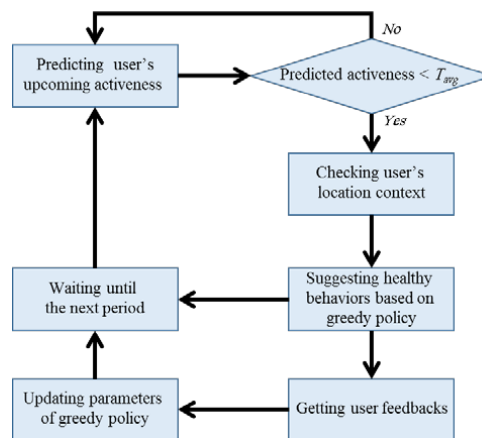


Figure 5 – Flow chart for suggesting appropriate behaviors to the user and adapting user feedbacks

Implementation of mobile wellness management system

In the previous two sub-sections, we explained the main steps of the wellness management system, RNN-based lifestyle modeling, and healthy behavior recommendation. Now, we describe the implementation details of our mobile app on Android platform. Figure 6 shows user interface snapshots for each functionality. The important functional features of the mobile app are as follows:

- Inquiry about the history of user's activeness for today or a specific date.
 - The user's activeness consists of heart rate, step count, and calorie consumption.
 - Each data of the activeness is depicted by a line chart individually.
- Recommend a healthy behavior periodically that can increase the user's wellness.
- Manage the history of activity suggestions.
- Support to get the user's feedback corresponding to activity suggestions.
 - There are two types of user feedback: a level of satisfaction regarding the prior suggestion, and another indicating whether a suggested activity is done or not.

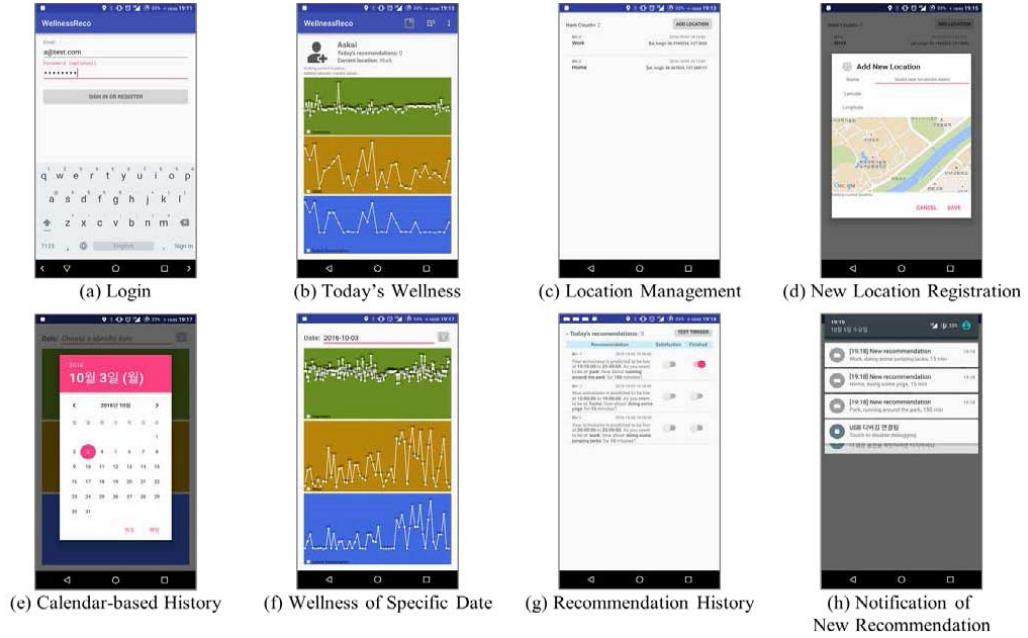


Figure 6 – Snapshots of user interface for the mobile wellness management app on Android platform

Experimental Design

We have plans to collect target users' lifestyle data by using Jawbone UP3 wristbands as shown in Figure 1. We need sufficient log of data for participants over a period of long time, at least for a few months. For logging precise data from each participant in our experiments, we asked that they follow the instructions below:

- Participants have to wear the wearable device whenever possible during the experiment.
- Participants have to answer a questionnaire of psychometric assessment tools every day.
- Participants have to give two types of feedback for each recommendation—a *satisfaction level* and an *execution rate* of suggested activities.
- When adding a new activity manually, participants have to honestly insert actual information.

The satisfaction level consists of 6-point Likert scale as follows: *strongly satisfied* (3), *satisfied* (2), *slightly satisfied* (1), *slightly dissatisfied* (-1), *dissatisfied* (-2), and *strongly dissatisfied* (-3).

The execution rate R_e is obtained by

$$R_e = \frac{\sum_{i=1}^N e_i}{N} \times 100$$

time, at least for a few months. For logging precise data from each participant in our experiment, where N is the total number of recommendations, e_i means the i -th flag of execution, and e_i is 1 if the suggested activity is done, 0 otherwise.

The results of questionnaire and feedback is used to verify if our quality measure properly works.

Conclusion

Maintaining an individual's QoL is necessary in modern society, but it is very difficult due to the variety of lifestyles. In this paper, we proposed a mobile wellness management system for

enhancing lifestyle quality based on analysis of time series data that is collected from wearable sensors. User's lifestyle patterns over time are extracted, and the RNN model is constructed by using the lifestyle features. Finally, our system makes suggestions of healthy behaviors underlying the learned RNN model depending on particular user's lifestyle.

In future work, we will conduct long-term experiment with several participants for collecting their activity logs of daily living and analyzing their lifestyle patterns in order to verify whether our management system has enough usefulness.

Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research for the Acquisition and Utilization of Field Expert Knowledge, funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

References

- [1] P.C. Austin, J.V. Tu, J.E. Ho, D. Levy, and D.S. Lee, Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes, *Journal of clinical epidemiology* **66** (2013), 398-407.
- [2] J. Bakker, M. Pechenizkiy, and N. Sidorova, What's your current stress level? Detection of stress patterns from GSR sensor data, in: *2011 IEEE 11th International Conference on Data Mining Workshops*, IEEE, 2011, pp. 573-580.
- [3] L. Bouarfia, L. Atallah, R.M. Kwasnicki, C. Pettitt, G. Frost, and G.-Z. Yang, Predicting free-living energy expenditure using a miniaturized ear-worn sensor: An evaluation against doubly labeled water, *IEEE Transactions on Biomedical Engineering* **61** (2014), 566-575.
- [4] T.M. Cheng, A.V. Savkin, B.G. Celler, S.W. Su, and L. Wang, Nonlinear modeling and control of human heartrate response during exercise with various work load intensities, *IEEE Transactions on Biomedical Engineering* **55** (2008), 2499-2508.
- [5] R.K. Dishman, Y. Nakamura, M.E. Garcia, R.W. Thompson, A.L. Dunn, and S.N. Blair, Heart rate variability, trait anxiety, and perceived stress among physically fit men and women, *International Journal of Psychophysiology* **37** (2000), 121-133.
- [6] R.I. García-Betances, G. Fico, D. Salvi, M. Ottaviano, and M.T. Arredondo, On the convergence of affective and persuasive

- technologies in computer-mediated health-care systems, (2015).
- [7] J. Hernandez, R.R. Morris, and R.W. Picard, Call center stress recognition with person-specific models, in: *International Conference on Affective Computing and Intelligent Interaction*, Springer, 2011, pp. 125-134.
 - [8] Z.C. Lipton, D.C. Kale, C. Elkan, and R. Wetzell, Learning to diagnose with LSTM recurrent neural networks, *arXiv preprint arXiv:1511.03677* (2015).
 - [9] A. Pande, Y. Zeng, A.K. Das, P. Mohapatra, S. Miyamoto, A. Seto, E.K. Henricson, and J.J. Han, Energy expenditure estimation with smartphone body sensors, in: *Proceedings of the 8th International Conference on Body Area Networks*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013, pp. 8-14.
 - [10] R.R. Rubin, Diabetes and quality of life, *Diabetes Spectrum* **13** (2000), 21.
 - [11] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks* **61** (2015), 85-117.
 - [12] C. Setz, B. Amrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, Discriminating stress from cognitive load using a wearable EDA device, *IEEE Transactions on information technology in biomedicine* **14** (2010), 410-417.
 - [13] J. Speight, M. Reaney, and K. Barnard, Not all roads lead to Rome—a review of quality of life measurement in adults with diabetes, *Diabetic Medicine* **26** (2009), 315- 327.
 - [14] T.G. Vrijkotte, L.J. Van Doornen, and E.J. De Geus, Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability, *Hypertension* **35** (2000), 880-886.
 - [15] H.A. Wolpert and B.J. Anderson, Management of diabetes: are doctors framing the benefits from the wrong perspective?, *British Medical Journal* **323** (2001), 994.

Address for correspondence

Chae-Gyun Lim (School of Computing, KAIST) is the corresponding author. E-mail: rayote@kaist.ac.kr.

Guidance Through Use: *Value* as a Pathfinder in e-Health Services Implementation

Bertil Lindenfalk^a, Vivian Vimarlund^{a, b}

^a International Business School, Jönköping University, Jönköping, Sweden

^b Department of Computer and Information Science, Linköping University, Linköping, Sweden

Abstract

The lack of awareness and confidence in eHealth solutions among certain stakeholders creates a barrier for the implementation of e-Health services. The aim of this paper is to explore issues that promote the development and implementation of patient-centered care services for the elderly. An exploratory case study approach is applied to a e-Health monitoring service that was developed and piloted in 38 homes for the elderly in Sweden and the Netherlands. The unit of analysis, concept of 'value-in-use', was used in order to determine how pilot participants felt about a service of this kind benefiting them the most. The findings were then translated into actionable considerations for implementing organizations. The results indicate a need for active participation, technical support infrastructure, mobility demands, and an extension of the concept of trust in e-Health services. The knowledge presented in the study is important for decisions makers, public organization strategists, and policy writers.

Keywords:

Health Services for the Aged, Humans, Medical Informatics

Introduction

E-health is a priority for the European Union (EU) and its member states [1]. Current research on the generic implementation of e-health services focuses on two things: (1) E-Health megatrends and its effects on health and social care for services in smart homes [2], and (2) removing existing barriers, such as the lack of interoperability among e-Health applications, limited large-scale evidence of the cost-effectiveness of e-Health tools and services, high start-up costs, regional differences in accessing ICT services, lack of legal clarity for health and wellbeing mobile applications, and the lack of awareness and confidence in e-Health solutions among stakeholders [1].

These challenges, in parallel with the growing number of the elderly, and the inability for many member states to give the elderly timely attention and support at the point of care have prompted EU member states to prioritize e-Health services that offer services with equal (or more) quality and effectiveness. At the same time, the re-allocation of resources and their proper use from an organizational perspective has to be allowed. This is especially the case in the member states that have publicly- or partially-funded health and social care systems.

Research in the areas of service marketing, management and service innovation, highlight the importance of the proper use and implementation of services as the fundamental basis of value creation, where exchange is but one component, along with use and context [3-5] issues are important to both private

and public organizations [3; 5-10]. This exchange cannot happen without the participation of both a service provider and a service beneficiary to co-create value. The service provider and the service beneficiary are, and always will be, interconnected [3; 4]. The goal of exchange is to use the applied knowledge of others (service) as resources to better one's circumstances. In the exchange between service systems (the provider and the beneficiary), value is determined through the use (or integration) and application of operant (and sometimes operand) resources [11]. The relationship is however, not equally dichotomous. The service beneficiary should always be considered as the most important part in the relationship. The service beneficiary decides the ultimate and final value of a service. Services are not of a purely monetary value and products are mere vessels for service distribution. All in all, all economies are service economies and value lies in use [3, 4, 7, 12]. For public sector organizations that govern developed e-Health services, it is also important to note that the value of e-Health services can be derived to society in addition to the individual service beneficiary [8]. This issue is something that current research in the field has not extensively reported on.

The aim of this paper is to explore issues that promote the development and implementation of patient-centred care services for the elderly, i.e., e-Health services, thereby addressing the barrier that "*the lack of awareness and confidence in e-Health solutions among stakeholders*" creates. The results of this study should be of interest for decision-makers, as well as policy writers.

The context of the study

The study was conducted in conjunction with the Healthy Life support through Comprehensive Tracking of individual and Environmental Behaviors project (HELICOPTER). The overall aim of the project was to support end-users' health in an unobtrusive and simple way, through monitoring daily-life behaviors, and supporting end-users and their caregivers with feedback, advice, and motivation in their pursuit of a healthy and safe lifestyle. The project was directed towards people over 65 who were not suffering from any major chronic diseases or severe disabilities, but who might possibly be affected by (or be at risk of) metabolic or circulatory malfunctioning (e.g., hypertension, mild diabetes), or by mild cognitive deficits.

The piloted service consisted of placing sensors in the homes of the pilot participants. In the finalized service, (see Figure 1 for a detailed overview of the service), the users are meant to be able to choose which sensors they would like to be incorporated into their own service experience, but for the purposes of the pilot all of the sensors were installed in the participating households. The sensors included a sensor placed

in the participant's favorite chair/couch, one bed-sensor for sleep-pattern recognition, one toilet sensor, one wearable sensor for monitoring the participant's movement, a sensor acting as a 'nose' in the refrigerator to check whether any food had gone bad, and sensors for drawers, etc. The participants also received a small computer to manage the data that was sent by sensors, a tablet to access the Helicopter service application, and what the project called a 'snowflake', which was supposed to act as an unobtrusive way of letting the user know whether the sensors were working or whether they needed to check the application for messages coming from the service. The tablet application was meant to act as an interface for signing, ordering the sensors at the beginning of the service lifetime and the interface where the service beneficiary could see long-term behavioral patterns, according to the information sent by the sensors.

In some of the households in the Netherlands, two clinical sensors were also included in the pilot. These were intended to establish a more concrete link to the healthcare practitioners and the healthcare system. This was intended to allow GPs who were part of the project to view the data that was collected by both the behavioral sensors and the clinical sensors, in order to investigate whether or not potential diseases could be inferred by the system. In the case of the project, focus was indicators of the early on-set of diabetes. However, due to certain technical and ethical considerations, the data collected by the clinical data sensors never actually left the sensors, i.e., no one but the person using the sensors could see what the sensors had monitored. This, of course, affected the general service process and hence the service experience.

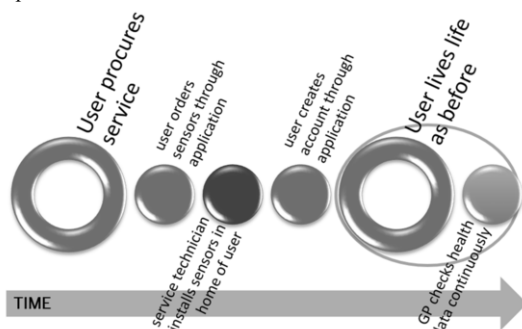


Figure 1 – General overview of the service process from the service beneficiary's point of view.

Framework for the study

e-Services have been defined as any service which to some extent uses the internet for service provision [13]. Typically, four different dimensions of importance are considered in relation to judging an e-Service's 'value-in-use'; these are the technical, functional, temporal, and spatial dimensions [14]. The technical dimension refers to the outcome of the service interaction, whilst the functional dimension encompasses the users' perception of the outcome of said service interaction. The temporal dimension refers to the user's perception of the time when the service interaction occurred and finally, the spatial dimension concerns the users' perception of the location where the service interaction occurs [14]. All of these dimensions play a key role in determining the service that is experienced from the beneficiary's point-of-view. Additionally, one of the antecedents of 'acceptance', a common concept in implementation research is, *use*. *Use*, is for instance, identified to be an important antecedent in implementation models such as TAM [15], UTAUT [16, 17],

as well as in IS continuance research [18]. *Use* is becoming an even more integral factor, when success in contexts, for multiple actors are involved is to be determined [7]. In this study, we adopt the e-service 'value-in-use' framework [19].

Method

The study used an explorative case-study approach [20] along with a holistic case study design, where the perceived 'value-in-use' was the single unit of analysis. Data was triangulated by using material collected from workshops, interviews, and questionnaires [20]. All activities were performed with the pilot participants after the service had been running for at least three months in each respective household, i.e., at the end of the project's lifetime.

Data collection procedures

Two workshops were held during the pilot. One with the Dutch participants (n=7) and one with the Swedish participants (n=9). The aim of the workshop was to get a collective view of who they thought was the potential user of the service, what they thought was valuable (and not so valuable) with the service, who was liable for what in the service and finally, what their view was of what the service could be in the future.

The workshops had the same structure in both countries, except in the Netherlands, a translator was used by the facilitator. The workshop consisted of two tasks for the participants. The first focused on the positive and negative experiences of using the service. The second task focused on the future potential value of the service.

The semi-structured interview (n=12) and the questionnaire (n=22) focused on the pilot participants' perceived value of the e-service and what impact the e-service might have on their daily lives. In the construction of the interview guide and the questionnaire, the e-service value-based framework [14, 19, 21] was used so as to address all of the relevant dimensions. For instance, questions like, "*Where and what time of the day did the participants use the tablet application?*", were asked. In addition, questions were asked about how they used the service and what value they could see in using the service personally. It is important to note that the questionnaire was descriptive in nature and was not intended to be used for statistical analysis. All material was translated into English from the participants' respective languages.

The analysis was conducted using the e-Service value framework where the technical, functional, spatial, and temporal dimensions were identified. This was followed by the categorization of positive and negative value dimensions. The analysis was conducted using Nvivo software where all the material from the activities was analyzed together in accordance with the method of data triangulation [20]. It's also important to note that feedback from the participants in the pilot was given to the service as it was piloted, i.e., potential improvements of the service would need to be considered, as well, in future development phases. The results of the study are thus based on illustrative quotes that were given by the pilot respondents and the subsequent discussion.

The pilot studies were carried out in the municipality of Skövde in Sweden, and in the Eindhoven Region of the Netherlands. In total, 38 pilot studies were carried out in end-user's homes in order to test the system. Consequently, the total number of respondents for each activity was 38 (n-max). Table 1 shows a detailed overview of the number of respondents for each activity. In total, 10 pilot studies were

carried out in Sweden (six single households and two couple households), and a total of 28 studies in the Netherlands (21 end users in 16 households) were included in September 2015, with an additional seven end users (in six households) were included in December 2015).

Table 1– The number of respondents for each the evaluation activities, in order of when they were performed

	Workshop	Interview	Questionnaire	n-max
Netherlands	7	3	13	28
Sweden	9	9	9	10
Total	16	12	22	38

Results

The result section is divided into five sub-sections that correspond to the four dimensions of the e-Service ‘value-in-use’ framework that was applied during analysis [14, 19] and one additional dimension which was identified during the analysis.

Technical value-in-use

The technical value is determined by service beneficiary’s assessment of the outcome of the service interaction, i.e., whether the service promise was fulfilled or not. For this dimension, certain aspects of the service, such as being able to create an account in the application, finishing the ordering procedure of the sensors, changing the batteries in the sensors, and the installation of the sensors became important issues. These constituted well-defined service interactions according to the users. However, after the instalment of the sensors, it became more difficult for the pilot participants to determine when they were actually using the service and when they were not using the service. This was due to the service’s unobtrusive nature. They had to trust that the sensors were working; something which they did not like. As one participant in the Swedish workshop put it; “the value lies in knowing that I’m okay”.

Functional value-in-use

Functional value refers to the service beneficiary’s perception of a service outcome, i.e., if they found it valuable or not, and if the interaction was deemed to be value-diminishing or value-amplifying.

One value-amplifier mentioned by the pilot participants was the fact that they could access their own data on the tablet application. The data collected day-by-day was not of particular interest, but rather the longitudinal data that was collected over time, where patterns could be noticed, was of interest. As one of the participants said, “seeing that the patterns are the same is a confirmation for me that I am okay”.

A value-diminisher that was mentioned by the pilot participants was that they did not like the fact that they could not choose the technological medium with which they were to access the service interface. They felt that being forced to use the tablet was not always the optimal solution for them, because they did not like the tablet as such. Some participants wanted to be able to use the tablets they already owned, whilst others felt that this was a service which they wanted to use their computer for. Another value-diminisher was the technical malfunction during the pilot where the application could not retrieve the data that was collected by the sensors. This was something that caused the participants to respond in an unforgiving manner towards the service and some pilot

participants even went as far as leaving the pilot study because of this malfunction. However, those participants who could access the data that was collected by the sensors in the application found this to be very interesting in the long term. As one participant put it, “I don’t care if my behavior changes from day to day, but perhaps week by week or month by month”, thus indicating how often she opened the application to view her behavioral charts.

One major value-diminisher was the use of the wearable activity tracker and identifier. This was used not only to measure the number of steps taken by the participant, etc. but, also to identify which particular user, for instance, used the toilet in a household where two people participated in the pilot. Here the pilot participants responded that the tracker was easily forgotten in the morning since they had to charge it every night, and so they felt that the data collected was not accurate. Some participants who did not have any pockets in the clothes they were wearing (predominantly the female participants) created pouches or other solutions so as to have the tracker on them.

Spatial value-in-use

The spatial value dimension refers to the service beneficiaries’ perception of the milieu where the service interaction occurred. This does not only include the interface where the service is co-created, but also takes into account non-service parts, such as the context and other people who are involved in providing the service.

A major value-diminisher of the entire service, according to the pilot study participants, was that the sensors only worked inside their homes; their activity outside the home was not recorded. This state of affairs left them to doubt the accuracy of the activity charts and the data that was collected. One participant stated, “I could be out and about all day, shopping, picking up my grandkids and playing with them, and be very active but when I look at my chart, it says that I have been idle. Which is not true!”.

One value-amplifier that was identified was how the sensors blended seamlessly into the home. Almost all participants, when asked if they were bothered by the sensors, said that they do not even notice them anymore. They did notice them in the first week, but since then, they were considered to be a natural part of the home. When they had guests over to visit, some participants said that some guests had asked what the boxes were, but once they were explained, guests took no further notice of them. The only sensor that they did notice on a daily basis was the wearable sensor. Some remarks were made about the bed sensor too, that it could be made more comfortable.

Temporal value-in-use

The *temporal value dimension* refers to the time when the service interaction occurs; including time dimension when the service co-creation could be aborted mid-way and be resumed at a later stage, or if certain processes either took too long, or were felt to be too short.

With respect to the temporal value in the service, the respondents found it difficult to define the time when they used the service once the sensors were up and running. Due to the unobtrusiveness of the service, the only service interaction that actively took place, from the pilot participant’s point of view, was when they tried to access their data in the application.

In terms of installation and setting up the service, the initial answer from the participants was unanimous; this took way too long and was way too complicated. The time spent on this

activity needs to be reduced and the sensors need to be very easy to install; almost to the level of merely inserting batteries before use. It was clear, from their point of view, that a service that took this much time to set up and install was something that they would never procure without knowing that the technical support offered with the service was excellent.

Towards Social value-in-use

In the interview and workshops, a fifth dimension was identified. Due to the fact that the service was an e-Health service, i.e. it was a service aimed at capturing diseases early on, the pilot study participants were inclined to tolerate certain flaws, for the benefit of society. They claimed that the service could be important, for society as a whole, and that they used the service so that other citizens could receive more acute care should they need it. As one pilot participant from Sweden phrased it, "I don't believe that we will have the resources required to accommodate everyone who needs care in the future and if we can, we should, use technology to minimize the need. It is very good". This dimension was not sought after in the interview questions, but it did appear when the question of why one should use a service like this was asked.

Discussion

Certain technological advancements and a growing need for pre-emptive care services that are directed towards the elderly are two forces that are changing the context in which future health and social care services are to be delivered [2, 9, 10]. In Europe, much of the work to be done lies in empowering citizens, by actively involving them in the care process through the use of e-Health services. Historically, many of the *new* services that have been developed by utilizing IT are still only implemented inside the boundaries of the care organization, e.g. the hospital. The provision of Electronic Health Records (EHR) is one such service. Our knowledge of the implementation of e-Health services is founded on the premise that users can be guided, and sometimes even forced, to use newly-developed services. However, with the contextual change that the nature of future e-Health services creates, the need for generic implementation models for organizations that act outside of the physical boundaries of an organization (e.g. a hospital or clinic) has arisen.

This paper argues that, by adopting a service view and carefully considering the factors that influence the successful implementation of said service, new and valuable knowledge can be generated which can be used to promote the development and successful implementation of e-Health services. Repeatedly, the importance of using e-Health services *and* services in general has been emphasized [3, 5, 8, 10, 15, 17, 22]. Therefore, we have adopted a framework specifically developed for e-Services where the notion of 'value-in-use' is emphasized [14, 19, 21, 23]. The framework consists of four 'value-in-use' dimensions; the technical, functional, temporal, and spatial dimensions.

The findings of this study can be condensed and translated into the following four implementation guidelines. Given the technical 'value-in-use' dimension, the users experienced difficulty in determining what the actual value of the service was when it was being used. This can, of course, be explained in terms of the unobtrusive nature of the service and the lack of active participation in the co-creation of the service from the service beneficiary's point of view. Previous research has shown that 'activity' is an important factor when services are co-created [21]. In terms of functional 'value-in-use', technological ease is still a major factor for the elderly. The

need for a specific technical support line was highly anticipated. From a spatial 'value-in-use' perspective, the need for mobility was highlighted; all access to e-Health services should not be determined by the location in which one finds oneself. Temporal 'value-in-use' aspects, on the other hand, emphasize the longitudinal time perspective of the service, as well as, the need for establishing a trustworthy baseline for each user.

The above remarks can be translated into the following actionable guidelines for implementation organizations. Clearly communicate not only the value proposition for the service but also the level of activity for service provision by the service beneficiary. In terms of infrastructure, it should be noted that the need for technological maturity stretches outside pure technical dimensions and includes infrastructure for service beneficiary tech-support. In the design and development phases, the mobility aspects of the service, must be taken into consideration so that the user's location should not be limiting factors for the provision of the service. When policies supporting the development of new e-Health services that focus on pre-emptive care are formulated, it is important to consider (for the sake of the trustworthiness of the service) that the service be provided while the service beneficiary still feels healthy. Previous studies have indicated the need for trust in relation to e-Health services and online provision of health information [24, 25]. These findings highlight yet another aspect of the trust issue that is related to e-Health services in general.

In general, the findings resonated well with the e-service 'value-in-use' framework. However, a value dimension which the original e-Service 'value-in-use' framework did not highlight was the *social* 'value-in-use' dimension. Although the findings are preliminary, they indicate that for e-Health services in the public-sector context, service beneficiaries project their own usage value to a larger societal context as well. It certainly might be the case that this dimension is culturally bound, since the pilot participants live in 'welfare states' to one extent or another. However, it is a finding that requires further attention than that given in this paper. Suggestively by considering conducting a longitudinal case study where a fully developed e-Health service is in use.

This study has several limitations. This work only considers the 'value-in-use' aspects of the entire value concept, leaving the value-in-context and value-in-exchange, as described by [5] the unexplored. Future studies could, for instance, focus on translating the value aspects found connected to use to those of exchange by considering a cost-benefit analysis. However, notwithstanding these limitations, interesting results that can be translated into actionable guidelines for organizations that wish to implement e-Health services can be found.

Another limitation is connected to the explorative nature of the single case-study. Generalizability for the study lies on a theoretical level, rather than on an empirical level. This is the case because exploratory case studies are almost impossible to replicate to the fullest [20]. The logic of replication dictates that, for future cases, the use of the same e-service 'value-in-use' framework as applied in this study should yield similar results. However, initial results of the exploratory case-study indicate that valuable insight might be gained by the continued investigation into 'value-in-use' for implementation research, where the focus is directed towards services for citizens in general. Suggestively, future studies would also include people with different types of digital literacy.

Conclusions

This study explored issues that promote the development and implementation of patient-centered care services for the elderly. It addressed “the lack of awareness and confidence in eHealth solutions among stakeholders”, as identified by the European Commission [1]. This was done by means of an exploratory case study that focused on the ‘value-in-use’ that elderly participants in the AAL Helicopter project experienced while using the piloted service.

Findings indicate that the ‘value-in-use’ concept and specifically the e-Service ‘value-in-use’ framework [14; 19] could give valuable insight for implementing organizations of pre-emptive e-Health services guided towards the elderly.

Acknowledgements

This research was funded by the AAL Joint Programme and the Swedish Innovation Agency; VINNOVA. The authors also extend their gratitude towards the AAL-HELICOPTER project consortium and its partners for their valuable contributions.

References

- [1] E. Commission, eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century, in, European Commission, Brussels, Belgium, 2012.
- [2] B.C.o. Australia, Overview of megatrends in health and their implications for Australia, in, Business Council of Australia, 2015.
- [3] S.L. Vargo and R.F. Lusch, Service-dominant logic: continuing the evolution, *Journal of the Academy of Marketing Science* **36** (2008), 1-10.
- [4] S.L. Vargo and R.F. Lusch, Evolving to a New Dominant Logic for Marketing, *Journal of Marketing* **68** (2004), 1-17.
- [5] S.L. Vargo, P.P. Maglio, and M.A. Akaka, On value and value co-creation: A service systems and service logic perspective, *European Management Journal* **26** (2008), 145-152.
- [6] S.L. Vargo and R.F. Lusch, It's all B2B...and beyond: Toward a systems perspective of the market, *Industrial Marketing Management* **40** (2011), 181-187.
- [7] S.L. Vargo and R.F. Lusch, From repeat patronage to value co-creation in service ecosystems: A transcending conceptualization of relationship, *Journal of Business Market Management* **4** (2010), 169-179.
- [8] J. Alford, Co-Production, Interdependence and Publicness: Extending public service-dominant logic, *Public Management Review* **18** (2016), 673-691.
- [9] S.P. Osborne, Z. Radnor, and G. Nasi, A New Theory for Public Service Management? Toward a (Public) Service-Dominant Approach, *The American Review of Public Administration* **43** (2013), 135-158.
- [10] S.P. Osborne and K. Strokosch, It takes Two to Tango? Understanding the Co-production of Public Services by Integrating the Services Management and Public Administration Perspectives, *British Journal of Management* **24** (2013), S31-S47.
- [11] R.F. Lusch and S.L. Vargo, Service-dominant logic as a foundation for a general theory, *The service-dominant logic of marketing: Dialog, debate, and directions* **406** (2006).
- [12] R. Normann and R. Ramirez, From value chain to value constellation: designing interactive strategy, *Harv Bus Rev* **71** (1993), 65-77.
- [13] H. Surjadjaja, S. Ghosh, and J. Antony, Determining and assessing the determinants of e-service operations, *Managing Service Quality* **13** (2003), 39-53.
- [14] K. Heinonen, Reconceptualizing customer perceived value: the value of time and place, *Managing Service Quality* **14** (2004), 205-215.
- [15] F.D. Davis, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly* **13** (1989), 319-340.
- [16] V. Venkatesh and F.D. Davis, A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies, *Management Science* **46** (2000), 186-204.
- [17] V. Venkatesh, M.G. Morris, B.D. Gordon, and F.D. Davis, User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly* **27** (2003), 425-478.
- [18] A. Bhattacharjee, Understanding Information Systems Continuance: An Expectation-Confirmation Model, *MIS Quarterly* **25** (2001), 351-370.
- [19] K. Heinonen, Temporal and spatial e-service value, *International Journal of Service Industry Management* **17** (2006), 380-400.
- [20] R.K. Yin, *Case Study Research: Design and Methods*, SAGE Publications, 2014.
- [21] K. Heinonen, The influence of customer activity on e-service value-in-use, *International Journal of Electronic Business* **7** (2009), 190-214.
- [22] V. Venkatesh, J.Y.L. Thong, and X. Xu, Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology, *MIS Quarterly* **36** (2012), 157-178.
- [23] K. Heinonen and T. Strandvik, Monitoring value-in-use of e-service, *Journal of Service Management* **20** (2009), 33-51.
- [24] B.W. Hesse, D.E. Nelson, G.L. Kreps, and et al., Trust and sources of health information: The impact of the internet and its implications for health care providers: findings from the first health information national trends survey, *Archives of Internal Medicine* **165** (2005), 2618-2624.
- [25] E. Vodicka, R. Mejilla, S.G. Leveille, J.D. Ralston, J.D. Darer, T. Delbanco, J. Walker, and J.G. Elmore, Online Access to Doctors' Notes: Patient Concerns About Privacy, *J Med Internet Res* **15** (2013), e208.

Address for correspondence

Corresponding author: Bertil Lindenfalk bertil.lindenfalk@ju.se

Developing “Information Assistant”: A Smartphone Application to Meet the Personalized Information Needs of Women with Breast Cancer

Ying Liu^a, Zhaohui Geng^a, Fulei Wu^a, Changrong Yuan^a

^aSchool of Nursing, Second Military Medical University, China

Abstract

Breast cancer presents specific challenges both physically and psychologically to women. Women living with breast cancer frequently do not receive adequate information support and they urgently need professional assistance that help them to cope and adjust to challenges. This paper documents the process of developing a smartphone application (app) “Information Assistant” for women with breast cancer in China. First, individual interviews were used to explore and understand the real experience of breast cancer patients and their diverse information needs during different stages of diagnosis and treatment. Next, expert working group discussions and the Delphi technique, including breast cancer care physicians, nurses and software engineers, were used to draft the structure and contents of the m-health based information support program. Then, these feedbacks were used to develop “Information Assistant” app. This app has five modules: personalized information recommendation, category knowledge center, headline information browsing, newest information browsing and information searching.

Keywords:

Breast Cancer; Information Management; Smartphone

Introduction

Breast cancer is the most common cancer among women worldwide [1-2]. According to the Chinese National Central Cancer Registry, breast cancer is the most common cancer among urban women and the fourth most common cancer in rural areas [3], cases in China account for 12.2% of all newly diagnosed breast cancers and 9.6% of all deaths from breast cancer worldwide [4].

Women with breast cancer have a higher relative survival rate than other cancers which makes it possible for patients to live a prolonged life [5]. Breast cancer poses specific challenges both physiologically and psychologically to women. A significant proportion of women experienced physical and emotional difficulties within the first year following diagnosis and years later after treatment [6]. Under these circumstances, information support has been identified as being vital to helping women to cope their challenges [7]. Information support not only increases the client’s involvement in the decision-making process which induces greater satisfaction with treatment choices, but also reduces the feeling of uncertainty and helps decrease emotional anxiety [7-9]. Thus, with adequate information support, women could feel empowered to fight their cancer and improve their quality of life [10].

However, the information needs of women with breast cancer have become increasingly difficult to meet as the number of survivors grows and the workforce of breast cancer care physicians and nurses faces personnel shortages. Since there is limited availability of time and resources by clinicians, there is

a need to identify a new method to provide adequate information support and improve the quality and efficiency of follow-up care for breast cancer survivors.

Recently, the rapid expansion of health information technology provides such an opportunity. It has been shown that health information technology can play an important role in helping cancer patients to get high-quality information and become active participants in their care [11-12]. Smartphones with variety functionality such as touch screen technology, audio, video, internet access have become a ubiquitous technology in both the developed and developing world [13]. With the increasing usage and acceptance of smartphones, the smartphone applications (apps) have shown great potential to provide clients with information specific to their needs [14]. Our project is to develop a smartphone app to meet the personalized information needs of women with breast cancer in China. The purpose of this paper is to document the process of developing the smartphone app.

Methods

This is a multidisciplinary research program, our research team included three postgraduate students (Ying Liu, Zhaohui Geng and Fulei Wu) majored in Chronic Disease Caring & Cancer Nursing, nursing educationist (Changrong Yuan), breast cancer care physicians, breast cancer caring nurses and software engineers. The research team members worked collaboratively to design and conduct the research. The workflow of the multidisciplinary research and developing of the app is presented in Figure 1.

Data collection spanned from March 2015 to December 2016. First, interviews and follow-up study with 12 breast cancer patients were used to explore and understand the real experience of breast cancer patients and their diverse information needs during different stages of diagnosis and treatment. Next, the expert working group discussion including 7 experts from medical, nursing and software engineering was conducted. Experts added some information need contents of breast cancer patients from their professional point of view, and gave suggestions for developing an m-health based information support program. Then, a Delphi expert consultation was conducted, in which 20 clinical nursing experts participated and gave their opinions on the core part of the mobile information support program. Finally, expert consultations including 5 clinical medical and nursing experts and 3 software engineers were conducted. A researcher organized the multi-disciplinary cooperation of the experts to revise and finalize the mobile information support program, and then work with software engineers to develop the software.

The study was approved by the National Natural Science Foundation of China and the Second Military Medical

University. All participating patients provided written informed consent to participate in the study.

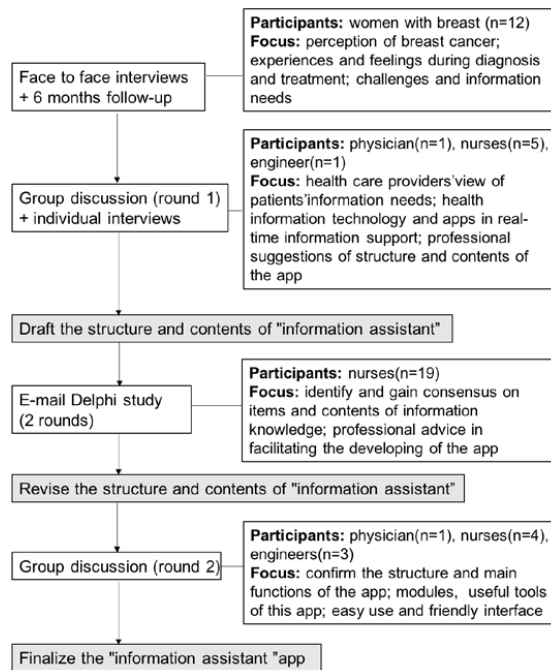


Figure 1 – Workflow of Developing "Information Assistant".

Interviews with Breast Cancer Patients

First, a qualitative approach was used to collect data about the information needs of women with breast cancer during diagnosis and treatment process. Twelve women who were diagnosed with breast cancer between March to April 2015 were selected to be interviewees. Eligibility for participating in this key informant interview included: (a) participants were diagnosed with breast cancer for first time, and receiving treatments; (b) participants were above 18 years old and they were able to communicate fluently in Chinese; (c) participants knew about their diagnosis and they were willing to talk about it; (d) participants were interested and willing to use smartphone apps to facilitate their breast cancer care. Informed consent was obtained from those participants.

We conducted at least two interviews with each patient and also had a 6 months follow-up to get the whole picture of patients' heterogeneous information needs during the different diagnosis and treatment process. The first interview for each woman was conducted during her diagnosis stage before the surgery, and the second interview was conducted about 2-4 days after her surgery. Some patients even had a third time interview just before they discharged from the hospital. The follow-up was started when the patient discharged from the hospital and went to his/her home or community for further rehabilitation, and all follow-up lasted for about 6 months.

All face to face interviews were conducted in conference rooms of several different hospitals in the east provinces of China and each interview lasted for 30-45 mins. The follow-up was conducted by telephone calls, short messages and the WeChat (a communication software).

Topics and questions related to patients' possible information needs were derived from the literature review and physicians'

and nurses' clinical experience. During the interviews and follow-up, participants were asked to describe: (a) their perception of breast cancer; (b) their experiences and feelings during breast cancer diagnosis and treatment process; (c) challenges they may face and information needs in living with breast cancer.

The individual interviews were audio-recorded with participants' consent and continued until data saturation was reached, meaning that no new information of value was obtained from the interviewees. Interviews were transcribed verbatim within two days of the interview.

A qualitative content analysis aimed at finding manifest and latent meanings of data was applied to analyze the information manually [15]. Firstly, the transcripts were read several times by the researcher, then segmentation of information was done and significant information related to research questions was extracted. Finally, data were coded and grouped into categories and abstracted into sub-themes and the main theme.

Groups Discussion Round One: Physicians, Nurses and Software Engineers

The group discussion with breast cancer care physicians, nurses and software engineers was conducted to draft the structure and contents of the "information assistant" app. The physicians and nurses had at least ten years working experience with breast cancer patients, and they were qualified with a masters degree or above, with the exception of one nurse who was qualified with a bachelor degree.

The discussions were conducted in a conference room at School of Nursing, the Second Military Medical University in China and lasted 120-150 mins. After the first round of group discussion, one-on-one interviews ranging from 20-45 minutes were conducted in a private room with those who were relatively active in the group discussion, to gain more suggestions.

The group discussion focused on: (a) what's the breast cancer patients' diverse information needs during the diagnosis and treatment process from health care providers' point of view; (b) what questions or problems patients consulted most in the follow-up caring of breast cancer; (c) how can health information technology and the app help in meeting patients' information needs and give real-time feedback?; (d) professional suggestions of structure and contents of the app.

Electronic Mail Delphi Study: Nurses

In order to identify and gain consensus on items and contents of information knowledge for addition into the app to support patients, and have more professional advice in facilitating the development of the app, we conducted an electronic mail Delphi study. The e-Delphi was more efficient and less time consuming. Nurses from all over China, who were not able to attend a face-to-face consensus conference or take part in nominal groups, were invited to participate in this study.

According to Keeney [16-17], Delphi does not use a random sample which is representative of the target population, rather, it employs 'experts' in the area in which the researcher is interested. As for this study, our research team defined and chose nurses who had more than five years working experience with breast cancer patients.

An introduction to the study and specific goals to reach were given to the participants in the first round consulting. The first round questionnaire consisted of two sections: the first section was the demographic information about the experts, including education background and working experiences. The second section was the main part. In this section experts were asked to give their opinions whether to accept or refuse or suggest for

revising about each item within the “eight information banks”. Some open-ended questions also included that experts can add more details on each information bank and give advice on revising the “information assistant” app.

Group Discussions Round Two: Physicians, Nurses and Software Engineers

We held the second round group discussion to finalize the “information assistant” app. Five experts in medical and nursing field who had already participated the first round group discussion were included, and the second round discussion also included three engineers. The discussion was conducted in the same conference room as previous and lasted between 120-150 minutes. We presented participants a protocol with revised structure and contents of the “information assistant” app. The discussion was focused on: (a) confirming the structure and main functions of the app; (b) modules, main functions and useful tools of this app; (c) easy to use and user-friendly interface; (d) the web-based administration portal.

Results

Draft the Structure and Contents of the “Information Assistant” App

Individual Interviews

In total, 12 women with breast cancer participated in the interview. The age of the participants ranged from 36 to 67 years, with a medium of 47 years. Half of them had a full-time job, eleven of them were married, and four had an education level above high school. As for their treatment, three of them had surgery only, seven had a combination of surgery and chemotherapy, and two had a combination of surgery and other adjuvant therapy.

Five major categories of information needs are identified as: (a) psychosocial support needs; (b) information about the treatment choosing and planning; (c) surgery process and related adverse reactions; (d) side effects of chemotherapy and coping strategy; (e) instructions of self-care at home. This part was reported in our previous work [18].

Group Discussion Round 1 and Individual Interviews

There were one physician (with 15 years working experience with breast cancer patients), five nurses (with an average of 18.8 years working experience with breast cancer patients) and one senior software engineer participated in the discussion.

During the discussion, we developed a draft of information support protocol. First, we provided the expert with the results of the qualitative study of the patients' true information needs, as well as our previous literature review findings in the group discussion.

Next, considering all patients' needs, professionals opinions and the breast cancer treatment guidelines [19], we identified eight key phrases in breast cancer caring, and these key phrases represented eight different stages in breast cancer care (see Table 1).

We assumed that patients in each stage can receive personalized information support through the app. Then we identified the concrete information contents in each stage; a total of 62 information items were identified. We called the overall stages and information items as “eight information banks”, and our next step would be to build the eight information banks.

Besides this, the group discussion also identified three more contents of the app as: (a) a category knowledge centre including all information; (b) a headline information browsing part for patients to get the most helpful information; (c) the

newest information browsing part for patients to get the latest information about breast cancer treatment and caring.

Revise the Structure and Contents of the “Information Assistant” App

There are 20 (a response rate of 91%) nursing experts participated in the round 1 e-mail Delphi study and they all completed the round 2 study. They came from five provinces (Shanghai, Zhejiang, Jiangsu, Hubei and Guangdong) and seven Grade III Class A hospitals in China. The age of the participants ranged from 29 to 54 years, with a mean of 38.9 years. The average working time was 17.0 years, ranging from 7 to 36 years. As for the education level, 15 of them had a bachelor degree or above.

The essential objective of this e-Delphi study was to access the suitability of each item and to identify any additional valuable information items in the eight information banks.

The overall original items of the eight information banks were 62. After the 2-round e-Delphi process, the experts reached consensus on 84-items information banks. According to the experts' suggestion, we added some useful information items in each bank so as to develop a more helpful app, enabling patients with breast cancer to get the maximum benefit.

The process of identifying information items during the 2-round Delphi study is shown in Table 1, and in Table 2 we present the contents of “adjuvant chemotherapy stage” as one example of the eight information banks.

Table 1 – Revised Information Items of Eight Information Banks.

Eight information banks	Original items	Delphi	Delphi
		1 round	2 round
Diagnosis stage	13	13	15
Adjuvant chemotherapy stage	3	6	8
Operation stage	13	16	17
Chemotherapy stage	8	10	11
Radiation therapy stage	4	5	5
Endocrine therapy stage	6	7	7
Targeted therapy stage	3	6	6
Rehabilitation stage	12	13	15
Total	62	76	84

Table 2 – Information Items of Adjuvant Chemotherapy Stage.

Item	Content of adjuvant chemotherapy stage
1	What's the meaning of adjuvant chemotherapy
2	What kind of people need adjuvant chemotherapy
3	Medications of adjuvant chemotherapy and it's indication
4	Choosing transfusion methods of chemotherapy
5	Caring of the peripherally inserted central catheter(PICC)
6	Caring of the implantable venous access port
7	Coping with the adverse effects of chemotherapy
8	Dietary and nutrition instruction during chemotherapy

Finalize the “Information Assistant”

In group discussion round 2, we went through our previous results, revised our contents and structure about the app and then finalized the app.

We did some modification of the app, for example, experts suggested that a searching bottom should be added to the interface of the app, to make it more user-friendly and easy use. Since most of our information support items were text materials, it may be difficult for some patients to catch up the

meaning. To facilitate the patients' understanding we added photo and video materials in the eight information banks, e.g., for breast self-examination and post-operation exercise.

Finally, an app with five main function modules was developed, as presented below:

Module One: Personalized Information Recommendation

This module is the major function of the app. In this module, we match the patients' present medical status, treatment planning and information preference with the "eight information banks". Based on this, we analyse and extract relevant information items from the bank, and then we deliver personalized information support to the patients.

Module Two: Category Knowledge Center

This module presents all the credible information that patients may want to know about breast cancer caring, including diagnosis, different tests, treatment regimens, medication management, dietary and nutrition instruction, and rehabilitation exercises instruction. Patients can find useful information in this module as complementary to the personalized information recommendation.

Module Three: Headline Information Browsing

This module is to provide the headline information which patients browse most. Patients best know what they want to learn. The top ten information items identified by real-time page views through our web-based administration portal could best represent issues that patients are concerned with.

Module Four: Newest Information Browsing

Both personalized information recommendation module and category knowledge center modules are based on the fixed eight information banks. This module is reserved to display latest research findings, new technologies and new methods about breast cancer caring, thereby helping patients get the most updated information.

Module five: Information searching

This module is also complementary to the personalized information recommendation. Patients who may have problems to find the information they need they can search in this module. Just like searching in google, patients can search by typing in the keywords and all credible information related within the eight information banks will be displayed on the screen.

Two screenshots are given as examples to visualize the smartphone app in Appendix (labeled as Appendix Figure 1 and Appendix Figure 2). The software is in the debugging process currently, the practical application and effectiveness test of the software will be carried out in our future study.

Discussion

Hearing of the diagnosis of breast cancer or just suspected of breast cancer, can put heavy pressure on patients, let alone enormous challenges they may face related to the various of treatments or rehabilitation strategies. Information support plays a vital role in helping patients to cope with their condition, as proved in many studies [7-10], and the effectiveness of web-based information systems has also been reported [20-21]. So, our focus in this work was not whether the information support would be useful for breast cancer patients, but how to adopt a more efficient and effective way to provide them with optimal information support. Inspired by the idea of m-health [13], a smartphone app which has huge advantages over the traditional ways [22] was developed based on a scientific method.

The primary goal of this "information assistant" is to meet the personalized information need of breast cancer patients, and provide each patient with personalized information support

according to their different medical status, treatment process and special needs. Based on interviews with breast cancer patients, group discussion with physicians, nurses and engineers, and e-Delphi study, the protocol of this "information assistant" app was developed. A qualitative study is a credible way to explore patients' true needs, and an e-Delphi study is an efficient way to gain consensus by experts. Besides, multidisciplinary group discussion promotes better communication and understanding between experts. These all contribute a lot to develop the app and deliver high-quality information support for the patients.

There were also some limitations of this study; first of all, the participating patients in interviews and experts in group discussions were relatively few and the findings may not cover all patients' information needs in various conditions; secondly, although e-health technologies have thrived in health care industry, there is no guideline or generally accepted principles for our health provider to develop a smartphone app. Finally, future research is needed to examine the usability and effectiveness of the "information assistant" app, and to optimize the app accordingly.

Conclusion

In this study, we tried a new method (smartphone app) to provide personalized information support for patients with breast cancer. By combining the results with patients' interviews, experts group discussion, and e-Delphi studies, we developed the "eight information banks" and developed the "information assistant" app. This app has five modules; personalized information recommendation, category knowledge center, headline information browsing, newest information browsing and information searching. Each module has a unique function to support patients optimally. Meanwhile, this process also gives a good example for future studies focusing on the design of e-health smartphone apps.

Acknowledgements

This research was supported by the National Natural Science Foundation of China. The project number is 71473262. No competing financial interests exist.

C.Y. is the principal investigator of this research project who was responsible for the entire study design. Y.L., Z.G., and F.W. performed the research and drafted the manuscript.

References

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, Cancer Statistics, *Ca-a Cancer J Clin* **65** (2015), 5-29.
- [2] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, and D. Forman, Global Cancer Statistics, *Ca-a Cancer J Clin* **61** (2011), 69-90.
- [3] National Cancer Center and Disease Prevention and Control Bureau, Ministry of Health, *Chinese cancer registry annual report*, Beijing: Military Medical Sciences Press, 2015.
- [4] L. Fan, K. Strasser-Weippl, J.J. Li, J. St Louis, D.M. Finkelstein, K.D. Yu, W.Q. Chen, Z.M. Shao, P.E. Goss, Breast cancer in China, *Lancet Oncol* **15** (2014), e279-289.
- [5] K. Katanoda and T. Matsuda, Five-year relative survival rate of breast cancer in the USA, Europe and Japan, *Jpn J Clin Oncol* **44** (2014), 611.
- [6] P.A. Ganz, L. Kwan, and A.L. Stanton, Physical and psychosocial recovery in the year after primary treatment of breast cancer, *J Clin Oncol* **29** (2011), 1101-1109.
- [7] C. Kwok and K. White, Perceived information needs and social support of Chinese-Australian breast cancer survivors, *Support Care Cancer* **22** (2014), 2651-2659.
- [8] M. Cappiello, R.S. Cunningham, M.T. Knopf, and D. Erdos, Breast cancer survivors: information and support after treatment, *Clin Nurs Res* **16** (2007), 278-293.
- [9] K.A. O'Leary, C.A. Estabrooks, K. Olson K, and C. Cumming, Information acquisition for women facing surgical treatment for breast cancer:

influencing factors and selected outcomes, *Patient Educ Couns* **69** (2007), 5-19.

- [10] R.E. Gray, M. Fitch, M. Greenberg, A. Hampson, M. Doherty, and M. Labrecque, The information needs of well, longer-term survivors of breast cancer, *Patient Educ Couns* **33** (2008), 245-255.
- [11] P. Klasnja, A. Hartzler, C. Powell, and W. Pratt, Supporting cancer patients' unanchored health information management with mobile technology, *AMIA Symposium* **2011** (2011), 732-741.
- [12] J. Blamoutier, Consumer health information seeking on the Internet: the state of the art, *Health Education Research* **16** (2011), 671-692.
- [13] M.H. Mobasheri, M. Johnston, D. King, D. Leff, P. Thiruchelvam, and A. Darzi, Smartphone breast applications - what's the evidence? *Breast* **23** (2014), 683-689.
- [14] C.M. Ruland, R.M. Maffei, E. Borosund, A. Krahn, T. Andersen, and G.H. Grimsbo, Evaluation of different features of an eHealth application for personalized illness management support: Cancer patients' use and appraisal of usefulness, *Int J Med Inform* **82** (2013), 593-603.
- [15] U.H. Graneheim and B. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, *Nurs Educ Today* **24** (2004), 105-112.
- [16] S. Keeney, F. Hasson, and H. McKenna, *The Delphi Technique in Nursing and Health Research*, Wiley-Blackwell, an imprint of John Wiley & Sons Ltd, 2010.
- [17] S. Keeney, F. Hasson, and H.P. McKenna, A critical review of the Delphi technique as a research methodology for nursing, *Int J Nurs Stud* **38** (2001), 195-200.
- [18] Y. Liu, T.T. Zhao, and C.R. Yuan, Qualitative research on information needs of patients with breast cancer in different stages of diagnosis and treatment, *Nurs J Chin PLA* **33** (2016), 31-34.
- [19] W.J. Gradishar, B.O. Anderson, R. Balassanian, S.L. Blair, H.J. Burstein, A. Cyr, A.D. Elias, W.B. Farrar, A. Forero, S.H. Giordano, M. Goetz, L.J. Goldstein, C.A. Hudis, S.J. Isakoff, P.K. Marcom, I.A. Mayer, B. McCormick, M. Moran, S.A. Patel, L.J. Pierce, E.C. Reed, K.E. Salerno, L.S. Schwartzberg, K.L. Smith, M. Lou Smith, H. Soliman, G. Somlo, M. Telli, J.H. Ward, D.A. Shead, and R. Kumar, NCCN Clinical Practice Guidelines in Oncology: Breast Cancer. Version 2, *J Compr Canc Ne* **13** (2015), 448-475.
- [20] M Ekstedt, E. Borosund, I.K. Svenningsen, and C.M. Ruland, Reducing errors through a web-based self-management support system, *Studies in health technology and informatics* **201** (2014), 328-334.
- [21] A.E. Wheelock, M.A. Bock, E.L. Martin, J. Hwang, M.L. Ernest, H.S. Rugo, L.J. Esserman, and M.E. Melisko, SIS.NET: A randomized controlled trial evaluating a web-based system for symptom management after treatment of breast cancer, *Cancer* **121** (2015), 893-899.
- [22] S. Lubberding, C.F. van Uden-Kraan, E.A. Te Velde, P. Cuijpers, C.R. Leemans, and I.M. Verdonck-de Leeuw, Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors, *J Clin Nurs* **24** (2015), 1367-1379.

Address for correspondence

Corresponding author: Changrong Yuan

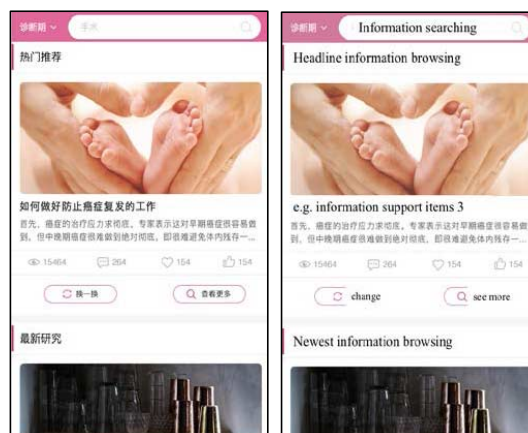
Email: yuancr510@aliyun.com.

Phone Number:021-81871490.

Appendix: Screenshots of "information assistant"



Appendix figure 1 – Left: main interface (Module one, two & five) ; right: english translation of main interface.



Appendix figure 2 – Left: main interface (Module three & four) ; right: english translation of main interface.

The Study of Smartphone Usage Competency Assessment and Training for the Elderly

Sheng-Chieh Lu^a, Tzu-Ning Wen^b, Po-Lun Chang^b

^a En Chu Kong Hospital, Taipei, Taiwan, China

^b Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

Abstract

This study aimed at developing an assessment of smartphone usage competence and constructing a training program for the elderly. A list of smartphone usage competencies with 34 items was defined through expert survey and panel. Based on the competence and previous literature, a training program and learning aids were designed in this study. There were 41 participants in our program. The results of self-administrated smartphone usage ability questionnaire indicated that all competencies were significantly improved after training. However, the results also demonstrated that some items were still difficult for the elderly to comprehend. Overall, this study provided a first exploration of defining smartphone usage competency and built a training program for the elderly. With strong suggestion, future mobile health (mHealth) services can follow this study to insure the smartphone usage ability of the elderly.

Keywords:

Aged, Education, Smartphone

Introduction

Noncommunicable diseases (NCDs) have become the leading cause of death around the world and caused 38 million deaths in 2012 [1]. Furthermore, higher prevalence of NCDs among older adults has been reported [2]. Due to irreversibility of NCDs, slowing down the process of the diseases and keeping healthy lifestyle have become the main targets of treatment. Under this situation, technology tools became important assistants for controlling the chronic diseases.

Mobile health (mHealth) represents a subset of telehealth. On account of rapid progress in mobile technology and widespread wireless network, mHealth became a novel model of health care service. Since the smartphone appeared, it has been quickly accepted by all ages around the world. Multisensory, wireless transmission and powerful computing power make smartphones a major mHealth device, and many smartphone-based health care services and research projects are ongoing [3,4].

Nevertheless, according to Aaron Smith's report, the smartphone adoption rate is relatively low in the aging population [5]. Many factors were identified as barriers for older adults to use smartphone, including decline of cognition, reduction of hand motor control, tardiness of information processing, and limited attention [7,8]. Furthermore, the elderly are rarely the main target of technology applications [9,10]. All above reasons are responsible for the lower smartphone adoption rate among the elderly.

In contrast, several researches have indicated that the obstacle of using smartphones for the elderly is not aging process or decline of learning ability but lack of proper training. If seniors can be provided with suitable training, they can develop greater ability of using information technologies and increase their willingness to use [9]. Therefore, the first task of serving the elderly with mHealth service is smartphone training programs.

Competency-based learning has been widely applied to design of training programs for workplace newcomers, medical students, as well as informatics literacy education [11]. The core value of competency-based learning is task-specific. Only highly task-related competence should be involved and all content of training program should base on the competence. Because the elderly do not need to be familiar with all functions and knowledge of the smartphone, the features of competency-based learning make it a strategy for developing training program.

Thus, the objectives of this study are: 1) define the essential smartphone usage competency for the elderly and 2) established a competency-based smartphone usage assessment and training program for the elderly.

Methods

The research divided into three parts: 1) define smartphone usage competence, 2) develop a smartphone usage training program and learning aids, and 3) apply the training program and evaluate the outcomes.

Define smartphone usage competence

Online expert surveys were applied to build the initial competencies. The questionnaire was designed with two open-ended questions to investigate what functions and knowledge of smartphones are needed for the elderly in mHealth. There were five experts surveyed, including three senior researchers in mHealth and two mobile application programmers. All functions and knowledge from the expert survey were analyzed by the research team and listed out to make up the initial smartphone usage competence for the elderly.

To validate the competencies, three senior researchers in health science of aging populations and one scholar who is a master in mobile healthcare system engineering were invited to constitute an expert panel. In the panel, each item of the initial smartphone usage competencies was reviewed for validity and readability. According to experts' opinions, smartphone usage competencies were proposed and used in constructing the smartphone usage training program for the elderly.

Develop smartphone usage training program and learning aids

Based on the competencies developed in previous stages, a smartphone usage training program was constructed for older adults. Each item in the competence list was put into the program. Acknowledging previous studies, our program follows four principles as stated below [10,12]:

- Include essential context
- Manage with task and scenario practice orientations
- Arrange the course according to the relations between each unit
- Design for levels ranging from easy to hard

The learning preferences of seniors were considered. The training strategy followed the principles as described below [6,12,13]:

- Allow learners with time to practice in the class
- Avoid using technical terms in the lecture
- The interval between sessions should not be long
- Provide several class dates for the elderly to choose

The learning aids were also designed following the suggestions of previous research. Previous studies have several recommendations for mobile phone instructions for seniors, such as [14]:

- Provide pictorial instructions, using words for aid
- Maximize the figures and words
- Make sure all contents are covered in the instructions
- Allocate all controls that will be used
- Start from the same view and never miss a step

Apply the training program and evaluate outcomes

We recruited a convenience sample of older adults from a metropolitan hospital in Taipei, Taiwan. We posted the information for our study and accepted applications from interested participants. Prospective participants were eligible to participate in the study if he/she was: more than 50 years old, and not hospitalized. All prospective participants were informed of the research procedure before recruited.

After being recruited, all participants completed a questionnaire that inquired about their demographic information (e.g. age, gender, education background, career) and their smartphone usage ability. The smartphone usage ability questionnaire was based on our smartphone usage competence and used a 5-point Likert scales to evaluate each item in the competence list. A competence score of 5 represented excellent competence, and 1 represented very poor competence. Next, each participant received an android phone and learning aids. All participants were asked to complete the whole training program. We offered 2 different time slot for each class, and all participants were asked to attend each class for one time. After all courses had completed, the elderly were required to self report their smartphone usage ability by the same questionnaire as beginning. The pre-test and post-test score were analyzed as the outcome of our training program. This study was approved by the Taipei City Hospital Institutional Review Board.

Results

The smartphone usage competency

There are ten smartphone functions and seven smartphone concepts that were proposed by the experts. The common reported-functions were calling (n=4), accessing internet (n=4), and system preferences setting (n=4). Other essential smartphone functions included SMS messaging (n=3), switching the phone into on, off, or standby modes (n=2), download applications from Google Play or other platforms (n=2), manage contacts (n=1), switching GPS on or off (n=1), and operating the touch keyboard (n=1). The frequently reported knowledge of smartphones included meanings of gestures (n=3), concept of network data flow (n=2), awareness of smartphone applications (n=2), conceptual understanding of the smartphone exterior and the position and function of each button (n=2), the difference between 3G and Wi-Fi (n=1), the meanings of icons on the notify bar (n=1), and conceptual understanding of the android home screen (n=1). By analyzing and organizing the sub-functions and related knowledge of 17 items, we listed out 41 core competencies of smartphone usage.

An expert panel was conducted to validate the smartphone usage competence. The suggestions of experts were 1) divide some items that contain two concepts or functions, 2) combine some similar items into a single item, 3) eliminate redundant items, 4) and simply and clarify the description of items to make it plain for the elderly. Besides, experts recommended that application downloaded through a text message should be added into the list. Referring these suggestions, 34 items were included in the final competence of smartphone usage and categorized in 7 groups (Table 1).

Smartphone usage training program

Based on the competencies, a smartphone usage training course and a series of learning aids were developed. The training course contained 12 units to cover all smartphone usage competencies. In order to shorten the course time, the course was separated into two classes and each class was no longer than 1 hour. The sequential order of the content in each class were based on the difficulty in each unit. Therefore, the first class began with introducing the fundamental knowledge and functions of smartphone, containing 1) introducing the appearance of smartphone and how to install the SIM card and battery, 2) switch smartphone to on, off, or standby mode, 3) turn the volume up or down, 4) answer and make a phone call, 5) maintain the address book, 6) send and receive SMS messages, 7) manage the home screen, 8) and take a photo and record video. The second class was more advanced. The topics were: 1) switch 3G, Wi-Fi and GPS on or off, 2) introduce the concept of network data flow, 3) introduce Google Play and how to download an app, 4) and download applications from a SMS message. All participants must attend each class and finished all courses in two weeks. The classes were lecture-based and the lecturer performed each of function step-by-step with the participants. After the demonstration, plenty of scenarios were given to participants and all participants were required successfully solve problems in these scenarios.

Smartphone usage learning aids

The learning aids contained two parts: a learning instruction with colorful comic strip (Figure 1) and a text-based manual booklet. The colorful instruction included all smartphone functions that were introduced in our training program. All functions were shown step by step and all essential controllers were pointed out on the instructions. In order to make the

instructions easier to read for older adults, the features of instruction were listed as follows: 1) all functions start from the same screen, 2) mainly show using pictures and assist with simple sentence, and 3) text size was 18px and pictures were as big as the smartphone used in this study.

Second, the text-based manual booklet focused on introducing the essential knowledge of using the smartphone (e.g. the functions of physical buttons, the meanings of icons, the concept of home screen, and concept of a network). Since the concepts were hard present by picture, the manual booklet mainly consisted of 14px text.

Table 1 - Smartphone usage competencies

Items
1. Fundamental knowledge of smartphone system
1.1 Install SIM card and Battery
1.2 Switch on the smartphone
1.3 Turn smartphone into standby mode
1.4 Wake smartphone from standby mode
1.5 Know when the battery need to be charged
1.6 Know how to charge the battery
1.7 Know the meanings of touch, press and hold, and drag when using the smartphone
1.8 Know the position and function of power, volume, menu, home and back button.
1.9 Know the position and function of microphone, speaker, back camera, and USB slot.
1.10 Know the meaning of icons (e.g. status of internet, battery, GPS, and reception) on the notify bar
1.11 Know the meanings of Standby, GPS, Google Play, and Wi-Fi
1.12 Able to discern buttons and text fields
1.13 Able to turn the volume up and down
1.14 Able to use touch keyboard to input data
2. Calling
2.1 Able to enter the phone number and make a call
2.2 Able to find a contact person from the address book and make a call
2.3 Able to make a call from communication records
2.4 Able to answer or hang up the phone
3. SMS messaging
3.1 Able to send and receive a short message
3.2 Able to download a application from a short message
4. Contact management
4.1 Able to add a new contact
4.2 Able to maintain contact information
4.3 Able to delete a contact
5. Take photo and record video
5.1 Able to take a photograph
5.2 Able to record a video
5.3 Able to browse the photo and video on the smartphone
5.4 Able to delete the photo and video
6. Manage home screen
6.1 Able to create and remove a shortcut of application
6.2 Able to create and remove a widget
6.3 Able to operate the widgets
7. Access internet and download applications
7.1 Able to switch the 3G wireless on and off
7.2 Able to search and download applications in Google Play
7.3 Able to switch GPS on and off
7.4 Know the meanings of network data flow



Figure 1 - Colored comic strip learning instruction

Implementation and evaluation of the smartphone usage training program

A total of 49 subjects participated in the study during July 2012 to November 2012, and 41 subjects completed the training program and self-reported questionnaire of smartphone usage ability. The age of participants were mostly between 60 to 69 years (65.9%), and were 39% male and 61% female. Participants reported variation of smartphone usage experience: 60.5% had not used smartphone, 12.2% had used smartphone less than one year, 19.5% had 1-3 years of smartphone experience, 2.4% had 3-5 years of smartphone experience, and 10.5% did not reply the question. The characteristics of the subjects and their previous experience of using smartphones is demonstrated in Table 2.

Pretest of smartphone usage ability

Overall results are shown in Table 3. The average pre-test score of smartphone usage ability was 1.91 ± 1.39 . Most participants evaluated themselves with poor or very poor in using smartphone. Participants reported higher score in calling (2.18 ± 1.67), contact management (2.05 ± 1.54), and fundamental knowledge of the smartphone (2.02 ± 1.47). In contrast, accessing the Internet and downloading applications (1.54 ± 1.08), managing the home screen (1.68 ± 1.18), and SMS messaging (1.91 ± 1.34) were reported with lower score.

Table 2 - Characteristics of study sample (N=41)

Variable	Frequency	%
Gender		
Male	16	39%
Female	25	61%
Age		
50-59 years	7	17.1%
60-69 years	27	65.9%
70-79 years	6	14.6%
Over 80 years	1	2.4%
Education		
Illiterate	1	2.4%
Primary	2	4.9%
Secondary	19	46.3%
Undergraduate	14	34.1%
Graduate	5	12.2%
Smartphone experience		
None	23	60.5%
Less than one year	5	12.2%
1-3 years	8	19.5%
3-5 years	1	2.4%
No reply	4	10.5%

Post-test of smartphone usage ability

After training, the average score of smartphone usage ability among participants was 3.16 ± 1.00 . Participants were inclined to score their smartphone usage as average or good. Calling, fundamental knowledge of the smartphone, and contact management were still evaluated with higher score (3.70 ± 0.87 , 3.64 ± 0.74 , 3.15 ± 1.08). Manage the home screen, accessing the Internet and downloading applications, and SMS messaging were relatively lower (2.76 ± 1.10 , 2.90 ± 0.96 , 2.89 ± 1.12). Comparing to the pre-test scores, all scores significantly raised ($P < .0001$).

Table 3 - Self-administed smartphone usage ability

Item	Pre-test	Post-test	t-test	
	Mean \pm SD	Mean \pm SD	t	p-value
Item1	2.02 \pm 1.47	3.64 \pm 0.73	-7.93	0.000*
Item2	2.18 \pm 1.66	3.71 \pm 0.87	-6.32	0.000*
Item3	1.81 \pm 1.34	2.89 \pm 1.12	-4.45	0.000*
Item4	2.05 \pm 1.54	3.15 \pm 1.08	-4.77	0.000*
Item5	1.96 \pm 1.48	3.10 \pm 1.15	-5.19	0.000*
Item6	1.68 \pm 1.18	2.76 \pm 1.10	-5.79	0.000*
Item7	1.54 \pm 1.07	2.90 \pm 0.96	-8.11	0.000*

* $P < .0001$; Item1-7 are referenced to table 2.

Discussion

Our study proposed a list of smartphone usage competencies and a training program to promote smartphone usage abilities of the elderly. Previous research has focused on training strategies, learning outcomes, and technology usage pattern of older adults [10,15,16]. Few studies, however, have emphasized the content of training programs and its advantages for elderly to receive mobile health care services. To our knowledge, our study is the first to investigate the smartphone usage competency needs and build a competency assessment, as well as training model for the elderly by ascertaining expert opinions.

The prerequisite for the elderly to participate in mHealth services is smartphone usage ability. The benefits of the services will be greater when seniors can master with all smartphone functions used in a certain service. Our smartphone usage competencies contain all functions that were frequently utilized in mobile health care services. According to the report by Free et al. (2013), calling, SMS messaging, GPS tracking, and downloading applications were the most widely used functions in mHealth services [4]. This demonstrates that our smartphone usage competencies can significantly match the needs required for mHealth services.

The contents of our smartphone usage training program were based on our smartphone usage competency. Therefore, it can be confirmed that well-designed programs can promote smartphone usage ability among the elderly. However, our program did not consider what the elderly is actually interested in. They are particularly interested in social applications include Line, Facebook and other entertainment applications. Although their needs may have no relation with the mHealth service, some previous studies have demonstrated that it will increase the motivation for adopting and learning among seniors when the technology fits their needs [17]. Thus, it is advised to moderately involve applications that interest the elderly, because these applications will motivate them to learn.

Our smartphone usage ability pre-test data show that most older adults evaluate their smartphone usage ability as poor or very poor. This finding is consistent with past research, which states that older adults often grade them self with low self-efficacy of technology use [6,18]. It also shows that the importance of smartphone usage training for the elderly before receiving the mHealth services.

After training program, post-test data illustrates that the usage ability of all functions included in our competence list were significantly improved. This finding provides with the insight that our training program and learning aids can positively promote the smartphone usage of older adults. Our findings also indicate that older adults still have ability to learn technologies, and this outcome is consistent with several previous reports [10,15,16].

Results also demonstrate that some specific functions and knowledge still are difficult for older adults to understand after training. Managing the home screen, accessing the Internet and downloading applications, and SMS messaging were the functions that were reported with lower scores. The possible explanation of the result could be an unfriendly interface design for the elderly. Although many studies have been started to give priority to interface design for the elderly, a senior friendly smartphone has not yet been produced [9]. Moreover, it is hard for the elderly to comprehend the concept of specific functions and knowledge because they have not had prior experience. Therefore, our study suggests that future programs for smartphone usage training among older adults should put more efforts on these functions and concepts.

There are several limitations needed to be considered in our study. Since our participants were convenience samples selected from a metropolitan hospital in Taipei, Taiwan, the majority of them live in the same region. The demographic features (i.e., education, economy, health status, and technology experience) of our samples may be dissimilar from older adults who live in another region. This aspect may limit our findings' generalizability. Additionally, it is difficult to highlight that all improvements of smartphone usage ability among the subjects were due to our training program since our study did not include a comparison group. Finally, past studies have indicated that a small class is better for the elderly [10]. However, under the consideration of teaching consistency and research resources, large classes (maximum 20 participants at once) were conducted for our study. The effect of class size on our outcome of training program was not examined.

Conclusion

Our study defines smartphone usage competencies for older adults and applies the competency-based learning as a strategy to build a smartphone usage training program. According to expert suggestions, our smartphone usage competencies cover all functions that have been frequently utilized in mobile health care services. The result of smartphone usage ability reported among our participants indicates that our training program can significantly improve the smartphone usage ability for older adults. With the ability to use a smartphone, the elderly can collect the health parameters through the smartphone and ensure the reliability of the data. We suggest that future smartphone-based health care services or studies can make reference to our research to make sure the data collected from the elderly are meaningful.

Acknowledgements

This research project was funded by Ministry of Science and Technology (PC10210-0020). The authors would like to thank all participants and anyone who contributed to this study.

Email address: polun@ym.edu.tw

Phone: 886-2-2826-7238

References

- [1] World Health Organization, Global Status Report on Noncommunicable Diseases 2014, World Health Organization, Geneva, 2015.
- [2] F. Wu, Y. Guo, P. Kowal, Y. Jiang, M. Yu, X. Li, Y. Zheng, and J. Xu, Prevalence of Major Chronic Conditions among Older Chinese Adults: The Study on Global AGEing and Adult Health (SAGE) Wave 1, *PLOS ONE* 8 (2013), e74176.
- [3] A.S.M. Mosa, I. Yoo, and L. Sheets, A Systematic Review of Healthcare Applications for Smartphones, *BMC Medical Informatics and Decision Making* 12 (2012), 67.
- [4] C. Free, G. Phillips, L. Galli, L. Watson, L. Felix, P. Edwards, V. Patel, and A. Haines, The Effectiveness of Mobile-Health Technology-Based Health Behaviour Change or Disease Management Interventions for Health Care Consumers: A Systematic Review, *PLoS Medicine* 10 (2013), e1001362.
- [5] A. Smith, Older Adults and Technology Use in, Pew Research Center, 2014.
- [6] S.J. Czaja, N. Charness, A.D. Fisk, C. Hertzog, S.N. Nair, W.A. Rogers, and J. Sharit, Factors Predicting the Use of Technology: Findings From the Center for Research and Education on Aging and Technology Enhancement (CREATE), *Psychology and Aging* 21 (2006), 333-352.
- [7] N. Charness and W.R. Boot, Aging and Information Technology Use: Potential and Barriers, *Current Directions in Psychological Science* 18 (2009).
- [8] Y. H. Wu, S. Damnée, H. Kerhervé, C. Ware, and A. S. Rigaud, Bridging the digital divide in older adults: a study from an initiative to inform older adults about new technologies, *Clinical Interventions in Aging* 10 (2015), 193-201.
- [9] S.J. Parker, S. Jessel, J.E. Richardson, and M.C. Reid, Older adults are mobile too! Identifying the barriers and facilitators to older adults' use of mHealth for pain management, *BMC Geriatrics* 13 (2013), 43-43.
- [10] B. Xie and J.M. Bugg, Public library computer training for older adults to access high-quality Internet health information, *Library & information science research* 31 (2009), 155-155.
- [11] R. Harris, H. Guthrie, R.B. Hobart, and D. Lundberg, Competency-based education and training: Between a rock and a whirlpool, Macmillan Education, Victoria, 1995.
- [12] C.B. Mayhorn, A.J. Stronge, A.C. McLaughlin, and W.A. Rogers, Older Adults, Computer Training, and the Systems Approach: A Formula for Success, *Educational Gerontology* 30 (2004), 185-203.
- [13] M. S. Knowles, E. F. Holton III, and R. A. Swanson, The adult learner: The definitive classic in adult education and human resource development, Routledge, London, New York, 2015.
- [14] C. Bruder, H. Wandke, and L. Blessing, Improving mobile phone instruction manuals for seniors, *Gerontechnology*; Vol 5, No 1 (2006).
- [15] A.A. Sterns, Curriculum Design and Program to Train Older Adults to Use Personal Digital Assistants, *The Gerontologist* 45 (2005), 828-834.
- [16] R.J. Campbell and D.A. Nolfi, Teaching Elderly Adults to Use the Internet to Access Health Care Information: Before-After Study, *Journal of Medical Internet Research* 7 (2005), e19.
- [17] M. Conci, F. Pianesi, and M. Zancanaro, Useful, Social and Enjoyable: Mobile Phone Adoption by Older People, in: 12th IFIP TC 13 International Conference, Sweden, 2009.
- [18] J.E. Chung, N. Park, H. Wang, J. Fulk, and M. McLaughlin, Age differences in perceptions of online community participation among non-users: An extension of the Technology Acceptance Model, *Computers in Human Behavior* 26 (2010), 1674-1684.

Address for correspondence

Po-Lun Chang, PhD.

Health Smart Homes: New Challenges

Anthony J. Maeder^a and Patricia A.H. Williams^b

^a School of Health Sciences, Flinders University, Adelaide, South Australia, Australia

^b School of Computer Science, Engineering & Mathematics, Flinders University, Adelaide, South Australia, Australia

Abstract

Health Smart Homes provide various forms of assisted living support, by monitoring the activities and health status of their occupants to generate flows of information and sometimes interventions involving the occupants and their careers. Technologies of varying complexity must be combined to produce the underlying Health Smart Home system, and processing of the resulting data may require methods of varying sophistication. These aspects have been well studied, but no widely-adopted approaches for practical implementation of systems or systematic processing of data have been developed. Also, the integration of Health Smart Home services with the overall health care system has not been regularized. This paper identifies and categorizes the emerging high-level challenges beyond those in the basic technical and algorithmic spaces. These challenges will influence future directions for Health Smart Homes and their wider adoption and integration with health systems.

Keywords:

Assisted Living Facilities; Computing Methodologies.

Introduction

Contemporary Smart Home concepts have been developed over the last two decades. A Smart Home consists of a living space equipped with electronic devices and communications infrastructure to enable monitoring and control of the immediate physical environment, coupled with a system with sufficient computational power, data storage and decision making capabilities to provide customization and efficiency of operation [1]. The emergence of Wireless Sensor Networks [2] and more recently, the Internet of Things (IoT) [3] and Cloud Computing [4] services, have provided the ideal technical fabric for constructing Smart Homes. Integration approaches relying on middleware and open source conversion platforms for various device and communications standards are now being supplanted by evolving Cloud-IoT synergies [5].

Health Smart Homes have been popularized alongside the mainstream Smart Home movement [6], with additional distinctive characteristics. Their purpose is more critical, in that they must measure, understand, and manage the health status of their occupants [7]. Early technical challenges identified for Health Smart Homes included the lack of appropriate sensor availability and performance [8], and difficulty in coordination of subsystems to provide component functions targeted to health needs [9]. The importance of social and ethical considerations for home occupants and subjects of care were also identified [10], and the need for multidisciplinary collaboration in the development of more

comprehensive health solutions [11]. Examination of Health Smart Home experimental implementations has revealed a wide spread of technology choices across numerous contexts [12], but little evaluation of their clinical effectiveness has been performed [13]. Despite attempts to standardize health sensor device protocols, such as the Continua framework [14], the lack of integration at a systems level remains a major limitation [15].

More recently, emphasis is being placed within health systems in the development of new models of care where consumers can receive a range of health services in new settings including their homes [16]. There is a clear consumer- supported move towards the widespread use of personal monitoring devices [17] that will need to be integrated into this emerging care environment. There is also a more general societal move towards wider access and use of personal electronic health records, and the implied connectivity of information systems, which will be required to achieve this [18]. Increasingly, interoperability and standardization are becoming aspirations for health systems that will ultimately rely on universal electronic data exchange and rich information environments supporting personalized “precision” patient care [19]. Health Smart Homes will need to harmonize with developments in these areas and, importantly, become part of a health services continuum, rather than isolated entities.

Methods

Given this broader perspective, we can readily distinguish two major challenge areas, Systems and Data, that will need to be addressed in the Health Smart Home of the future. In the following sections, we identify aspects within these two areas that are currently open issues and for which substantial work is needed to make further progress.

Systems Challenges

In the context of whole-of-systems concepts for Health Smart Homes, we comment on three new challenge areas: Architecture issues, including overall design considerations; Integration issues, such as interfacing and interoperability; and Safety issues, involving protection of users and the systems themselves.

Architecture Issues: The computing-based components consisting of: (i) hardware units; (ii) connecting networks; and (iii) intelligent software modules, which make up a Health Smart Home, can be viewed as a Critical System, in the same sense as large scale physical infrastructure systems (e.g., power grid) or complex machinery control systems (e.g., passenger aircraft). Such systems need mechanisms to allow update and replacement of components without disrupting system function, and addition or deletion of components without compromising

the integrity and performance of the system. This would imply an ability to assure the effects of any such changes, and to predict their impact, on other components should be embedded in the system development methodology [20]. Consistency of protocols for managing data and control within critical systems, including movement towards self-validation to provide error resilience to changes [21] along with convergence of Wireless Sensor Networks with IoT and Cloud Computing technologies [22], will help advance this cause.

Integration Issues: To achieve fuller knowledge of past health history and to conduct longer time frame modeling of occupants' complex health circumstances (e.g., when monitoring a subject with multiple chronic diseases), it may be necessary to connect with external sources of health information on the individual and leverage sophisticated external health decision making and services delivery systems. This implies interoperation of the Health Smart Home with other health services will be necessary, which in turn will have associated impacts on workflows and care models. Some reconceptualization of the Health Smart Home as one of the essential components in the overall health system would help to address this matter. This would include differentiating between the role of supporting healthy living (e.g., preventive health coaching and lifestyle monitoring) [23] and that of supporting existing acute health care delivery (e.g., the numerous "hospital in the home" initiatives) [24]. Monitoring to support acute care at home is currently arguably more socially acceptable and the intrusiveness of the required technology tolerated if not accepted. However, in designing for non-acute care and supporting everyday wellbeing, this level of acceptance and related engagement diminishes [25].

Safety Issues: As wireless connectivity becomes ubiquitous individuals are placed at higher risk of accidental or malicious interference with electronically managed data streams and control functions. This security vulnerability is amplified in high density Wireless Sensor Networks and IoT settings [26] such as Smart Homes, and where the sensitivity of health data to breach must be considered. Protection against such situations necessitates rigorous enforcement of failure detection and recovery, audits and monitoring to determine misuse, and system risk management processes to mitigate against failures. Robustness and resilience components in the system design cycle need to be developed, and operational surveillance protocols need to be established so that warnings can be generated and fail-safe defaults adopted. Whether it is in the pre-determined sensor networks or the healthcare IoT environment, the fail-safe modes of devices, such as power failure incidents, must be carefully considered. The fail modes - default on or default off - need to be contextually considered. For example, under some conditions in falls monitoring, a false negative (e.g., from a slow fall) may be more damaging than a false positive (e.g., bending down to reach the floor). Further, the device susceptibility to attacks and the challenge of applying traditional security models with delineated network boundaries to dynamic Smart Home and Healthcare IoT networks, pose significant problems which demand more attention in system standards development [27].

Data Challenges

Extracting more value from health data is currently a major emphasis in formulating major electronic health record and Clinical Decision Support strategies. This is beginning to extend to the use of data from sources such as Health Smart Homes monitoring. Three aspects of this trend contribute to the emergence of new issues here: Management of the related data-

sets including content and custodial issues; Recognition of relevant patterns in the data to allow accurate continuous classification of an individual's activities; and Personalization of decision making concerning health events or status, relative to the individual's aggregated data.

Data Management Issues: As in hospital-based patient record systems, strict establishment and adherence to standards for Health Smart Home datasets will be necessary, and equally, the interpretation of externally sourced health data on individuals will need to be informed by knowledge of those standards. Little existing standards work has been reported for Health Smart Home settings and this inhibits wider use and sharing of datasets with other elements in established health systems. If this was addressed it would become easier to provide reference sets of experimental data for algorithm tuning and benchmark data for performance analysis. In the same way as standards were adopted in the research community for sharing physiological datasets via PhysioNet [28], a similar practice would be desirable for Health Smart Homes where sharing is much less common [29]. Another related matter is that of the processing of the data in a Health Smart Home to ensure clinical usefulness as well as appropriate detection and response to potential health indicators. The healthcare workflow and the integration of data from sensors with clinical knowledge is of vital importance to assure improved health outcomes [30].

Pattern Recognition Issues: Characterization and reliably detecting patterns associated with health conditions (e.g., using physiological signals) or behavioral manifestations (e.g., mobility tracking) remains an open problem. Much effort has been invested in attempts to create robust algorithms for such pattern extraction using machine learning and artificial intelligence methods on data which may be sparser than desired due to physical limitations in the collecting infrastructure [31]. The highly variable nature of repeated situations and activities, both between and within subject, and allowing for "drift" in patterns over time and in different contexts can confound the success of these approaches [32]. Statistical detection of anomalies and structuring patterns into a hierarchy of macro to micro pattern characterizations offer some positive directions for longer term adaptation within activity algorithms, leading to greater consistency of results [33]. The enormity of the range and tolerance issues associated with achieving high sensitivity and specificity performance for detecting and classifying daily living activities, will likely maintain this challenge for research into the foreseeable future.

Subject Personalization Issues: Personalized application of algorithms in a Health Smart Home is an essential feature of its operation: if real health benefits are to be gained then it is necessary to tune the system as well as possible to fit the characteristics of the occupying individuals. This may require considerable sophistication in classifier decision-making logic, as well as dealing with multilayer redundancy through the existence of overlapping data sources (e.g., wearables vs ambient) in the data collection and processing stages [34]. Interference from interactions with other people and from externally imposed events which affect the individual, can degrade performance considerably. The more factual data that can be collected on an individual, and the more information that captures the influence of context and current situation, the more opportunity there is to create a fuller personalized picture [35]. Incorporating knowledge of phenotypical habits such as a weekly cycle of behavioral interactions and incorporation of indirectly measurable affective parameters such as mood and stress, may also contribute to overcoming these limitations [36].

Discussion

The simplicity of typical Health Smart Home solutions and the lack of compelling evidence for their clinical effectiveness has been identified as a major limitation in their adoption [37]. On the other hand, the increasing importance of home-based healthcare delivery and the consequent acceleration in the need for Health Smart Homes has been noted [38]. The clinical evidence for keeping people at home and avoiding unnecessary hospital admissions is plentiful [39]. Integrating social and health care, patient self-management, socio-economic status as well as coordination with primary care providers are some of the factors that can impact this [40]. The use of Health Smart Homes could be a vital component for improvement in keeping people at home longer.

It is probably unreasonable to expect that any of the new challenges identified above will be resolved by incremental progress, or by awaiting an unforeseen breakthrough followed by widespread adoption. The history of Smart Home development has been marked by independent engineering contributions and implementation models driven in a bottom-up manner [41]. A much broader, overarching approach will be needed to catalyze progress in these areas, as they will affect other aspects of the technology. For example, if a standard were to be agreed for data captured by sensors in a Health Smart Home, which prescribed that side information on precision and reliability must be included, then much previous work would be rendered obsolete.

Conclusions

The evolution of Health Smart Homes to achieve widespread presence requires more than technology advancement. The challenges identified above are subtle and complex consequences of the broader environment for health care within which the Health Smart Home functions. However, it seems likely that these challenges will influence future directions in Health Smart Homes, their wider adoption, and their integration with existing health systems. We advocate that more attention should be paid to these broader issues and that related research questions must be explored, alongside the current proliferation of underpinning engineering and implementation work.

Acknowledgements

Support of Cisco Systems Australia Pty Ltd and Flinders Digital Health Research Centre in the undertaking of this research is gratefully acknowledged.

References

- [1] S.S. Intille, Designing a home of the future, *IEEE Pervas Comput* **1** (2002), 76-82.
- [2] G.J. Pottie and W.J. Kaiser, Wireless integrated network sensors, *Commun ACM* **43** (2000), 51-58.
- [3] N. Gershenfeld, R. Krikorian and D. Cohen, Internet of Things, *Sci Am* **291** (2004), 76-81.
- [4] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg and I. Brandic, Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility, *Future Gener Comp Sy* **25** (2009), 599-616.
- [5] A. Botta, W. De Donato, V. Persico and A. Pescapé, Integration of cloud computing and internet of things: a survey, *Future Gener Comp Sy* **56** (2016), 684-700.
- [6] V. Rialle, F. Duchene, N. Noury, L. Bajolle and J. Demongeot, Health smart home: information technology for patients at home, *J Telemed e-Health* **8** (2002), 395-409.
- [7] I. Korhonen, J. Parkka and M. Van Gils, Health monitoring in the home of the future, *IEEE Eng Med Biol* **22** (2003), 66-73.
- [8] M. Chan, D. Estève, C. Escriba and E. Campo, A review of smart homes—Present state and future challenges, *Comput Meth and Prog Bio* **91** (2008), 55-81.
- [9] D. H. Stefanov, Z. Bien and W.C. Bang, The smart house for older persons and persons with physical disabilities: structure, technology arrangements, and perspectives, *IEEE T Neural Sys Reh* **12** (2004), 228-250.
- [10] G. Demiris, M. J. Rantz, M.A. Aud, K.D. Marek, H.W. Tyrer, M. Skubic and A.A. Hussam, Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study, *Med Inform Internet* **29** (2004), 87-94.
- [11] S. Koch, Healthy ageing supported by technology—a cross-disciplinary research challenge, *Inform Health Soc Ca* **35** (2010), 81-91.
- [12] S. Koch, M. Marschollek, K.H. Wolf, M. Plischke and R. Haux, On health-enabling and ambient-assistive technologies, *Meth Inf Med* **48** (2009), 29-37.
- [13] G. Demiris and B.K. Hensel, Technologies for an aging society: a systematic review of "smart home" applications, *Yearbook of Medical Informatics* **3** (2008) 33-40.
- [14] R. Carroll, R. Cnossen, M. Schnell and D. Simons, Continua: An interoperable personal healthcare ecosystem, *IEEE Pervas Comput* **6** (2007).
- [15] M. Memon, S.R. Wagner, C.F. Pedersen, F.H. Beevi and F.O. Hansen, Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes, *Sensors* **14** (2014), 4312-4341.
- [16] M. Chan, D. Estève, J.Y. Fourniols, C. Escriba and E. Campo, Smart wearable systems: Current status and future challenges, *Artif Int Me* **56** (2012), 137-156.
- [17] J. F. Coughlin and J. Pope, Innovations in health, well-ness, and aging-in-place, *IEEE Eng Med Biol* **27** (2008).
- [18] N. Archer, U. Fevrier-Thomas, C. Lokker, K.A. McKibbin and S. E. Straus, Personal health records: a scoping review, *J Am Med Inform Assoc* **18** (2011), 515-522.
- [19] W. E. Hammond, C. Bailey, P. Boucher, M. Spohr and P. Whitaker, Connecting information to improve health, *Health Affair* **29** (2010), 284-288.
- [20] K. Benghazi, M.V. Hurtado, M.J. Hornos, M.L. Rodríguez, C. Rodríguez-Domínguez, A.B. Pelegrina and M.J. Rodríguez-Fórtiz, Enabling correct design and formal analysis of Ambient Assisted Living systems, *J Syst Software* **85** (2012), 498-510.
- [21] M. Chen, J. Wan, S. González, X. Liao and V.C. Leung, A survey of recent developments in home M2M networks, *IEEE Commun Surv Tut* **16** (2014), 98-114.
- [22] J. Gubbi, R. Buyya, S. Marusic and M. Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, *Future Gener Comp Sys* **29**(7) (2013), 1645-1660.
- [23] J.A. Botia, A. Villa and J. Palma, Ambient assisted living system for in-home monitoring of healthy independent elders, *Expert Syst Appl* **39** (2012), 8136-8148.
- [24] G. Caplan, N. S. Sulaiman, D.A. Mangin, N. Aimonino Ricauda, A.D. Wilson and L. Barclay, A meta-analysis of "hospital in the home", *Med J Australia* **197** (2012), 512-519.
- [25] T. Irizarry, A.D. Dabbs and C.R. Curran, Patient portals and patient engagement: a state of the science review, *J Med Internet Res* **17**(2015).
- [26] R.H. Weber, Internet of Things—New security and privacy challenges, *Comput Law Secur Rev* **26** (2010), 23-30.
- [27] P.A.H. Williams and V. McCauley, Always Connected: The Security Challenges of the Healthcare Internet of Things, *Proceedings of the IEEE World Forum in Internet of Things* (2016), 30-35.
- [28] G.B. Moody, R.G. Mark and A.L. Goldberger, Physio-Net: a web-based resource for the study of physiologic signals, *IEEE Eng Med Bio* **20** (2001), 70-75.
- [29] A. Fleury, M. Vacher, F. Portet, P. Chahua and N. Noury, A multimodal corpus recorded in a health smart home, *Proceedings of the International Conference on Language Resources and Evaluation LREC 2010* (2010), 99-105.
- [30] P.A.H. Williams and A.J. Maeder, A conceptual framework for secure use of mobile health, *J Int Soc Telemed eHealth* **1** (2013), 44-51.
- [31] J. Ko, C. Lu, M. B. Srivastava, J.A. Stankovic, A. Terzis and M. Welsh, Wireless sensor networks for healthcare, *Proceedings of the IEEE* **98** (2010), 1947-1960.
- [32] L. Chen, J. Hoey, C.D. Nugent, D.J. Cook and Z. Yu, Sensor-based activity recognition, *IEEE T Syst Man Cy C* **42** (2012), 790-808.
- [33] P. Rashidi, D.J. Cook, L.B. Holder and M. Schmitter-Edgecombe, Discovering activities to recognize and track in a smart environment, *IEEE T Knowl Data En*, **23** (2011), 527-539.
- [34] S. Okour, A. Maeder and J. Basilakis, Multi-layered system design for classifying activities of daily living, *St Heal T* **178** (2012), 157-162.
- [35] A. Harvey, A. Brand, S.T. Holgate, L.V. Kristiansen, H. Lehrach, A. Palotie and B. Prainsack, The future of technologies for personalised medicine, *New Biotechnol* **29** (2012), 625-633.
- [36] R.A. Calvo and S. D'Mello, Affect detection: An interdisciplinary review of models, methods, and their applications, *IEEE T Affect Comput* **1** (2010), 18-37.

- [37] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz and A. R. Rincon, Smart homes and home health monitoring technologies for older adults: A systematic review. *International J Med Inform* **91**(2016), 44-59.
- [38] R. Haux, S. Koch, N.H. Lovell, M. Marschollek, N. Nakashima and K.H. Wolf, Health-Enabling and Ambient Assistive Technologies: Past, Present, Future. *IMIA Year- book of Medical Informatics* **11** (2016), S1-S16.
- [39] M.V. Williams, A requirement to reduce readmissions: take care of the patient, not just the disease, *J Am Med Inform Assoc* **309** (2013), 394-396.
- [40] S. Purdy, Avoiding hospital admissions: What does the research evidence say? *The King's Fund* (2010), Available from <https://www.kingsfund.org.uk/sites/files/kf/Avoiding-Hospital-Admissions-Sarah-Purdy-December2010.pdf>
- [41] M.R. Alam, M.B. Reaz and M.A. M Ali, A review of smart homes—past, present, and future, *IEEE T Syst Man Cy C* **42** (2012), 1190-1203.

Address for correspondence

Prof. Anthony Maeder, Flinders University, GPO Box 2100 Adelaide SA 5001, Australia

Email: anthony.maeder@flinders.edu.au

Telephone: +61882013107

Prof Trish Williams, Flinders University, GPO Box 2100 Adelaide SA 5001, Australia

Email: patricia.williams@flinders.edu.au

Telephone: +61882012023

Dynamic Creation of Patient Summaries: A CDA and IHE XDS Based Approach for Regional EHRs

Patrick Mangesius^a, Samrend Saboor^b, Thomas Healy^c, Thomas Schabetsberger^{a,d}

^a ITH icoserve GmbH, Innsbruck, Austria,

^b Siemens Healthineers GmbH, Erlangen, Germany

^c Siemens Healthineers GmbH, Digital Services, USA

^d University of Health Sciences, Medical Informatics and Technology, Hall, Austria

Abstract

Cooperative healthcare is regarded as one of the major goals for providing adequate health related information to physicians. To achieve this goal, national authorities and large hospital organizations are introducing large scale, standard based transactional EHR systems. Those systems record and distribute a significant amount of medical related data. This raises the concern of information overload for the intended users. The objective is to elaborate on an architecture and consequently a workflow that allows the generation of an automatic patient summary in a standard based IHE XDS environment. A literature review evaluating the current state of research is conducted. Current eHealth projects, laws and technical background are analyzed. An architecture is suggested, prototyped and compared using SAAM (Software Architecture Analysis Method) against alternative approaches. A technical workflow built on IHE XDR and HL7 FHIR observations is suggested introducing two new services within an IHE XDS product for extracting observations from CDA documents and storing the data on domain level scope. The information is published as an OnDemand Document in the IHE XDS infrastructure.

Keywords:

Electronic Health Record, Information Management

Introduction

For information systems and software products operating in the professional domain of eHealth, one major goal is providing health related information to physicians and patients adequately and timely [1]. Providing medical practitioners with comprehensive electronic health records (EHRs) and information is intended to contribute to an optimized treatment process and essentially improve continuity of care [2]. Therefore, this is seen as a prerequisite to a more efficient and cooperative healthcare [3]. The definition and quality aspects of EHR systems have been evaluated and published according to several approaches over the last years [4]. Subsequently, national states as well as the EU are focusing on introducing regional/nationwide and connected EHRs [5; 6]. As a current example, Austria introduced ELGA “Elektronische Gesundheitsakte” (ELGA) [7]. It is a nationwide, standards-based EHR infrastructure which went online officially in December of 2015. ELGA is regarded as a virtual EHR system for all Austrian citizens connecting all hospitals and a patient portal in several phases [5]. Due to their nature, such large scale EHR systems are expected to record and distribute more medical information and reports than conventional electronic medical records (EMRs) in isolated data silos. Benefiting from

this vast information overflow is a challenge for physicians and systems [8]. However, doctors’ acceptance of EHR systems is one major, critical success factor [9]. Presenting the patient’s vital information in a summarized view, extracted from the information stored in the EHR, could lead to a reduction in workload and an improvement in quality of care [10]. This set of data is referred to as a patient summary, a standardized set of medical data including facts that are necessary for a safe and efficient patient treatment. On a European level the eSOS project outlined a special form of such a patient summary [11]. This electronic patient summary is referred to as a “Consolidated Continuity of Care” (C-CDA) document. It has been defined by HL7 and required by the US-driven “Meaningful Use Stage 2” in the HITECH Act [12]. The objective of this paper is to elaborate on and conceptualize the ELGA implementation as an architectural validation of implemented IHE profiles to generate automatic patient summaries. The suggested approach respects currently established interoperability standards and workflows of large-scale EHR systems.

Methods

Focusing on recent work, a literature review of research on the extraction of data and creation of patient summaries was conducted. The research supporting this paper focused on recent publications found in Pubmed and the ACM Digital Library. Only documents published after 2003 were considered. For the identification of relevant papers, the following keywords and combinations were used: “ehr data summary”, “ehr data extraction”, “cda data extraction”, “cda patient summary”, “clinical document data extraction”. Relevant references from already included papers were also included. In addition to the analysis of relevant academic papers, current market leading standardization organizations and the technologies of current eHealth projects were identified. National data privacy and health telematics laws as a basis of regional/national EHRs [13] and the technical background of the implemented architectures were analyzed. The newest developments from the standardization units and organizations, whose technology and standard definitions were identified as being used in current eHealth projects, were analyzed.

Using this information, a technical concept intended to seamlessly integrate into a standardized eHealth architecture was developed. In order to validate the feasibility of the concept, a prototype was implemented and integrated into a demo installation of a major eHealth EHR product (sense® by ITH icoserve GmbH, a Siemens company). The product was selected because it is deployed in major parts of the Austrian

ELGA EHR and thus fulfills all requirements defined for the national project [7].

As a final step the architecture was compared with two other approaches discovered in the literature review executed in the initial step. The Software Architecture Analysis Method (SAAM) [14] was used for comparison. The compared architectures are described in [15] and [16].

Results

The following set of standards and technologies were identified as crucial:

- IHE IT Infrastructure Framework (IHE XDS, XDS-I, PIXPDQ, XDR, ATNA and related profiles)
- HL7 Clinical Document Architecture (HL7 CDA)
- HL7 V2, V3 and Fhir as Data- and Communication Model
- Logical Observation Identifiers Names and Codes (LOINC)

The literature review also confirmed the initial assumption concerning the need of a summarization of patient information stored in an EHR infrastructure. Electronic medical records (EMR) and EHR systems still primarily operate on the basis of documents as atomic information entities [17]. For users, this is in strong contrast to the requirements for the EHR to grant fast access to discrete data [18] as well as an adequate and scalable retrieval of information [19]. As the patient summary is intended to be part of an EHR system, this requirement is also mandatory for the proposed architecture. Another aspect is that EHR systems are required to strictly enforce a tight and well defined access control system limiting the user's rights and abilities. This also impacts the data included in a patient summary since users with different rights may have access to different data presented in the summary. The contained data is also dependent on the treatment context in which the user is accessing the EHR. We concluded that it is necessary to generate a patient summary on demand. In order to fulfill the previously mentioned requirement of the access performance, we suggest a system that uses a database of previously extracted information. This database is populated in the workflow of registering documents in the distributed infrastructure such as defined by IHE XDS [17].

In such an environment the registration workflow is executed by three services implementing the following IHE profiles: Document Source, XDS Repository and XDS Registry, as depicted in Figure 1. The system by definition does not analyze the documents content, the nature of the documents is reflected by the document metadata.

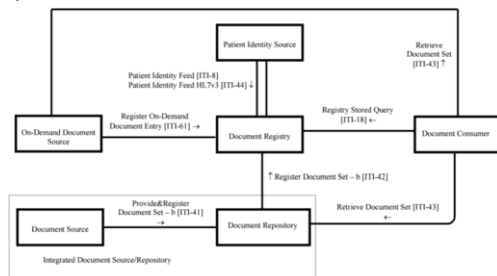


Figure 1– IHE XDS-b Architecture Overview [17]

To overcome this document based approach we introduce two new services: Observation Extraction Service (OES) and

Observation Broker Service (OBS). The OES is intended to operate within the scope of the local institution; the OBS is operating on scope of the IHE affinity domain. The deployment scenario and overview of the architecture is depicted in Figure 2.

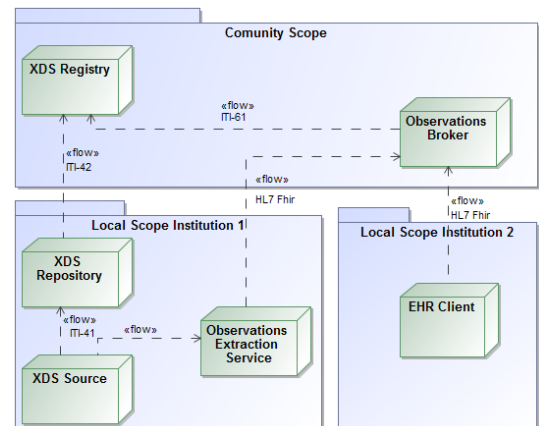


Figure 2– Deployment Overview of an Observations Enhanced IHE XDS Infrastructure

The main task of the OES is to extract data from the CDA documents intended to be registered in the EHR system. To assure standard based communication it is expected to provide an IHE XDR Recipient compliant interface [17], accepting the ITI-41 (Provide and Register Document Set-b) transaction. Using this interface, a standards based XDS Source can submit the data triggering a second ITI-41 transaction to the OES in addition to registering it in the infrastructure. Medical information is expected to be represented in a structured and coded format within the CDA documents, e.g. in the context of the Austrian national EHR ELGA data is coded using the LOINC (Logical Observation Identifiers Names and Codes) Standard. LOINC is a database which provides a universal code system for reporting laboratory and other clinical observations [20]. Using this information, the OES can extract discrete data. To make the data available to all authorized users and systems accessing the affinity domain the data is submitted to the OBS using a REST based HL7 FHIR transaction [21]. The OBS is intended to collect discrete data of the patient. These include, but are not limited to, the data extracted from the previously extracted documents. It is also possible, using the HL7 FHIR transactions to submit directly the discrete values that were collected in the process of progress monitoring of a patient, laboratory data, vital signs from home monitoring, or other [21]. For publishing the data to the affinity domain, the OBS registers an On-Demand Document to the affinity domain using the IHE Transaction ITI-61 [22]. This document, which is assembled at the time of access using the data registered in the OBS service, allows XDS based clients to retrieve the data wrapped in a CDA L3 document using the standard ITI-43. In addition, it is also possible to access the data using an HL7 Observations Query, if the client wants to avoid processing an XML based documents (e.g. on mobile devices, where REST based API calls are preferred [23]). The workflow of the registering and retrieval process is outlined in Figure 3.

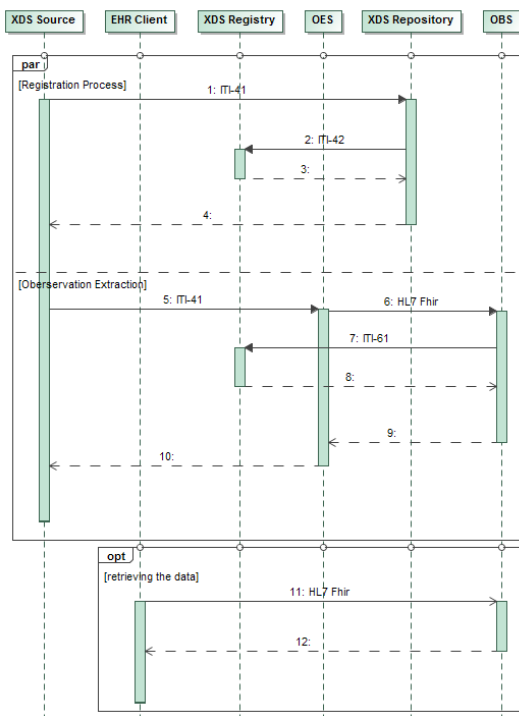


Figure 3—Workflow of the Registering and Retrieval of Observations extracted from CDA Documents.

Another aspect that needs to be considered is the management of data that is needed from the extraction process. Using this workflow, the OES receives the whole document metadata in the initiating ITI-41. Using this information, the service is capable of deciding which documents are relevant for the extraction process and which are not. Additionally, in the document analysis process, the service can decide which information is sent to the OBS and which is not. This, however, is likely to be defined by the scenario the services operate in. This avoids the unnecessary duplication of irrelevant data.

Discussion

The objective of this paper was to elaborate on a conceptual architecture on a technical level of how patient summaries can be generated on the fly in an IHE standard based infrastructure. Starting off with a literature review, the need for an architecture solving this problem could be confirmed. The literature review—although the queries were conducted using broad search terms—resulted in only three relevant approaches. An explanation for this circumstance might be that standard based XDS EHR systems just started to be widely established in the last years.

In contrast to the other approaches, the workflow suggested in this paper tries to integrate into the registering process of medical documents. Capturing the data in the phase of registering allows the extraction of concrete data from the documents without executing automated transactions that do not involve human actors. Automated non-user-based, transactions are always regarded as suspicious and hard to explain to the patient who will see the automated access in the audit logs of the EHR, a functionality of EHR systems that is crucial for patient acceptance and in many countries required by law. [5]. In addition, using this approach, quality requirements addressing performance and feasibility can also

be addressed. The direct extraction of the information from the EHR system would require the client to retrieve all relevant documents of the patient at runtime. Depending on the amount and size of the documents, this would take a significant amount of time and therefore result in negative impact on the usability of the solution. This drawback is avoided by designing a service containing pre-extracted and processed observation information.

The approach integrates seamlessly into standards-based EHR systems using the IHE XDS infrastructure. Using the directed submission functionality of the IHE XDR profile, documents can be directly sent to the service extracting the medical data from the document. This OES Service is within the scope of a local installation, e.g. of a hospital network infrastructure. The approach is also expected to scale well under high load since the extraction process itself is not carried out on a central infrastructure. The data itself are stored in a centralized service, the Observation Broker. This service is able to generate the patient summary itself. The output will either be an On-demand document [22] or a HL7 FHIR based query interface. The implemented prototype proves that it could seamlessly fit into a standard based IHE compliant product and does not require major additions to already existing workflows and standards as they are recommended in large eHealth projects like ELGA [24].

Despite these advantages, there has to be some additional work executed concerning the context and access rights. EHR systems, especially in the IHE XDS environment, are commonly built on an access control system focusing on restricting user access according to rules referencing the XDS document metadata, the role of the user, and the presence of a relationship or association between the user and the patient. As the referenced approaches within this architecture separate the discrete medical data from the document container, a concept needs to be elaborated, to efficiently establish a connection between the document metadata based access control system and the extracted discrete data segments. Another open topic is the auditing requirement of accessing the patient summary. While audit requirements for access to clinical documents is quite clearly defined by standardization units such as IHE or national laws such as ELGA, the audit specification for access to extracted discrete data elements or generated values are less clear. This concerns the pure information of the access to the information as well as what information is shown, when the data are accessed by a certain user.

Conclusion

Providing a summarized and aggregated view of the patients most essential information is an important feature for users of large scale EHR systems. To achieve this, discrete values to information must be provided in addition to achieving this in document-based EHR systems implementing the IHE XDS standard. This information can be extracted from the structured CDA document. Establishing this extraction process in the workflow of registering documents allows the processing of the documents without triggering automated, consuming side transactions, which are regarded as suspicious by patients and users. Additionally, extracting the information in the registration process has a positive effect on the performance, since the information is already present as discrete values and must not be gathered on the fly while the user is waiting for the information. It is, however, still an open challenge on how to restrict the access to the extracted information. Due to the document based nature of IHE XDS EHR systems the access control system of such deployments also work on a document

level. Solving this problem needs additional research and will be the focus of future work.

References

- [1] S. Lütkehaus, *Der Patient und seine Akte: elektronische Patientenakten und das Selbstbestimmungsrecht*, BoD–Books on Demand, 2010.
- [2] I. Iakovidis, Towards personal health record: current situation, obstacles and trends in implementation of electronic healthcare record in Europe, *International journal of medical informatics* **52** (1998), 105-115.
- [3] K. Stroetmann, T. Jones, A. Dobrew, and V. Stroetmann, The economic benefits of implemented eHealth solutions at ten European sites, *Bonn (Germany): eHealth-impact* (2006).
- [4] A. Hoerbst, T. Schabetsberger, W. Hackl, and E. Ammenwerth, Requirements regarding quality certification of Electronic Health Records, in: *Medical Informatics in a United and Healthy Europe: Proceedings of MIE 2009, the XXII International Congress of the European Federation for Medical Informatics*, IOS Press, 2009, p. 384.
- [5] BMG (Bundesministerium für Gesundheit), Elektronische Gesundheitsakte-Gesetz – ELGA-G, in: BMG, ed., 2012.
- [6] European Commission, Directive 2011/24/EU of the European Parliament and of the council of 9 March 2011 on the application of patients' rights in cross-border healthcare, *Official Journal of the European Union* **4.4.2011** (2011).
- [7] ELGA GmbH, *ELGA Gesamtarchitektur*, ELGA GmbH, 2015.
- [8] A.S. O'malley, J.M. Grossman, G.R. Cohen, N.M. Kemper, and H.H. Pham, Are electronic medical records helpful for care coordination? Experiences of physician practices, *Journal of general internal medicine* **25** (2010), 177-185.
- [9] K. Steininger, B. Stiglbauer, B. Baumgartner, and B. Engleder, Factors Explaining Physicians' Acceptance of Electronic Health Records, in: *2014 47th Hawaii International Conference on System Sciences*, IEEE, 2014, pp. 2768-2777.
- [10] EU Commission Directorate General Health & Consumers, Guidelines on minimum/nonexhaustive patient summary dataset for electronic exchange in accordance with the cross-border directive 2011/24/EU, (2013).
- [11] S.A.o.L.A.a. Regions, epSOS: Electronic patient smart open services, in, <http://www.epsos.eu>.
- [12] Department of Health and Human Services, Federal Register, *Federal Register*, **77** (2012).
- [13] M.L. time.lex, Overview of the national laws on electronic health records in the EU Member States and their interaction with the provision of cross-border eHealth services, (2014).
- [14] R. Kazman, G. Abowd, L. Bass, and P. Clements, Scenario-based analysis of software architecture, *IEEE software* **13** (1996), 47-55.
- [15] A. Schuler, O. Krauss, and B. Franz, Automated on-demand generation of patient summary documents, (2015).
- [16] C. Rinner, M. Kohler, S. Saboor, G. Huebner-Bloder, E. Ammenwerth, and G. Duftschmid, Searching for document contents in an IHE-XDS EHR architecture via archetype-based indexing of document types, *Studies in health technology and informatics* **192** (2012), 1092-1092.
- [17] Integrating the Healthcare Enterprise (IHE), Infrastructure Technical Framework, in: *Cross-Enterprise Clinical Documents Sharing (XDS), by the IHE ITI Technical Committee, Version*, 2015.
- [18] J.S. Ash and D.W. Bates, Factors and forces affecting EHR system adoption: report of a 2004 ACMI discussion, *Journal of the American Medical Informatics Association* **12** (2005), 8-12.
- [19] A. Hoerbst and E. Ammenwerth, Electronic health records, *Methods Inf Med* **49** (2010), 320-336.
- [20] C.J. McDonald, S.M. Huff, J.G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, and J. Hook, LOINC, a universal standard for identifying laboratory observations: a 5-year update, *Clinical chemistry* **49** (2003), 624-633.
- [21] HL7.org, HL7 Fhir: Resource Observation, in, HL7.org, HL7.org, 2016.
- [22] Integrating the Healthcare Enterprise (IHE), On-Demand Documents - Technical Framework Supplement, in: *On-Demand Documents - Technical Framework Supplement, by the IHE ITI Technical Committee, Version*, IHE, 2013.
- [23] K. Wagh and R. Thool, A comparative study of soap vs rest web services provisioning techniques for mobile host, *Journal of Information Engineering and Applications* **2** (2012), 12-16.
- [24] Bundesministerium für Gesundheit Familie und Jugend, Standards im österreichischen Gesundheitswesen, in, UBIT/BMGfJ - I/A/15, 2007.

Analysis of HL7 EHRS Functional Model and Suggested Applications in China

Haibin Meng, Huajian Mao, Chenyang Chi, Dongsheng Zhao

Institute of Health Service and Medical Information, Academy of Military Medical Sciences, Beijing, China

Abstract

HL7 electronic health record system (EHRS) functional model is an international standard which gives scientific instructions for EHRS design and implementation. It has the features of technological irrelevancy, requirement hierarchy and functional comprehensiveness. It is composed of overarching, care provision, care provision support, administrative support, population health support, record infrastructure and trusted infrastructure. This paper describes the functionalities of each section in detail. To apply this functional model, the history of Chinese health information system construction is reviewed and the weakness is analyzed as a basis. Then this paper gives suggestions on the construction of health information system in China: to enhance the functionality of healthcare systems, to enhance the information network function of regional health information network (RHIN), and to redesign the trust architecture of RHIN.

Keywords:

Electronic Health Records; Models, Theoretical; Suggestion

Introduction

An Electronic Health Record (EHR) is a repository of information regarding the health status of a subject of care in computer processable form. Electronic Health Record System (EHRS) is a set of components that form the mechanism by which electronic health records are created, used, stored and retrieved, including people, data, rules and procedures, processing and storage devices, and communication and support facilities [1]. EHRS functional model (EHRS FM) defines a standard functionality list that presents in an EHRS [2]. As a core model, it determines the main functionality framework that an EHRS should have. At present, China is on the process of constructing interoperable population health information networks [3]. It is essential to get theoretical instructions from the EHRS FM. In this paper, we introduce the history and principles of HL7 (Health Level 7) EHRS FM, and describes the main content of HL7 EHRS FM Release 2 (R2) in detail. Then we review the history of Chinese health information system construction and analyze the weakness of this construction mode. Further, we give suggestions on applying EHRS FM to improve the health information system construction in China.

HL7 EHRS FM

History of HL7 EHRS FM

The research of HL7 EHRS FM can be traced back to April, 2003. It was sponsored by HL7 workgroup. In July 2004, it was approved as a draft standard for trial use. The HL7 EHRS FM R1 was released in February 2007 and became the ANSI standard. The HL7 EHRS FM R1.1 was released in November

2009 and was accepted as an international standard ISO/HL7 10781 [4]. In April 2014, the HL7 EHRS FM R2 was released. Compared with the previous edition, this edition made fundamental change by proposing an entirely new architecture of EHRS FM. It became the ISO standard in 2015 [5].

Principles of HL7 EHRS FM

The principles of HL7 EHRS FM can be summarized as technologically irrelevancy, requirement hierarchy, and functional comprehensiveness.

Technologically Irrelevancy

With the development of information technology, new technologies emerge continuously. The functionalities of EHRS are constantly implemented by new technology, but the functionality framework of EHRS is relatively stable. Therefore, one principle of EHRS FM is to separate the functional framework from its implementation. The technologically irrelevancy is realized by defining the scope and content of EHRS functionalities instead of the technological details of EHRS.

Requirement Hierarchy

Owing to the diversification of EHRS, the actual EHRS conforms to a specific functional profile which is a subset of the EHRS functional model. The conformance criteria have two levels: the required criteria and optional criteria. The required criteria are expressed by mandatory word *shall*. The optional criteria can be divided into two sub-levels. Suggested criteria are expressed by word *should*. Permit criteria are expressed by word *may*. Requirement hierarchy bring about the flexibility of EHRS FM. It can meet the needs of various application situations and adapt to various development stages of EHRS.

Functional Comprehensiveness

Unlike the functional model of common information systems, HL7 EHRS FM greatly extends the scope of EHRS functionalities. Especially in the arena of information network architecture and trust architecture. It embodies the functional comprehensiveness of EHRS.

Description of HL7 EHRS FM R2

Figure 1 shows the general structure of HL7 EHRS FM R2. It consists of 7 sections, which are Overarching (OV), Care Provision (CP), Care Provision Support (CPS), Administrative Support (AS), Population Health Support (POP), Record Infrastructure (RI) and Trusted Infrastructure (TI). Each section contains functionality list described by ID, type, name, statement, description and conformance criteria.

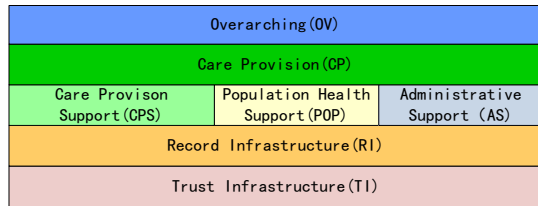


Figure 1–The General Structure of HL7 EHRS FM R2

Overarching (OV)

The Overarching section contains conformance criteria that apply to all EHRS. There are 2 functionalities and 33 conformance criteria in this section, including health record report and output, record lifecycle management, security management, terminology service and system management.

Care Provision (CP)

This section contains functionalities and conformance criteria required to provide direct care to a specific patient and practice of healthcare. Organized in general flow of an encounter, these functionalities can be used as EHR supporting functionalities to various healthcare systems. There are 41 functionalities and 494 conformance criteria organized in 9 categories, which are managing clinical history, rendering externally-sourced information, managing clinical documentation, managing orders, managing results, managing medication/immunization/treatment administration, managing future care, managing patient education & communication, and managing care coordination and reporting.

Care Provision Support (CPS)

This section contains functionalities that support the provision of care. It has 76 functionalities and 559 conformance criteria. They are classified into 10 categories as record management, supporting externally-sourced clinical documents, supporting clinical documentation, supporting orders, supporting for results, supporting treatment administration, supporting future care, supporting patient education and communication, supporting care coordination and reporting, and managing user help.

Administration Support (AS)

The Administration Support section includes functionalities that provide support for the management of the clinical practice and give assistance to the administrative and financial operations, such as resource management, workflow, and communication with patients and healthcare providers. It also provides support for managing non-clinical administrative information on patients and providers. It has 55 functionalities and 249 conformance criteria. They are categorized as provider information management, patient demographics and location management and synchronization, personal health record (PHR) interaction management, communication management, clinical workflow tasking management, resource availability management, encounter/episode of care management, information access for supplemental use, and administrative transaction management.

Population Health Support (POP)

This section includes functionalities for supporting public health management and disease prevention and control. It focuses on special group of people, such as chronic disease group, women, and children. These functionalities can support data aggregation for medical research, public health promotion, and improving the quality of care. Besides, this section also includes functionalities about protecting patient privacy and supporting patient consents for the secondary uses of medical records. There are 18 functionalities and 108 conformance

criteria in this section. They belong to 10 categories: support for health maintenance/preventative care/wellness, support for population-based epidemiological investigation, support for notification and response, support for monitoring response notifications regarding a specific patient's health, support for donor management, measurement/analysis/research and reports, public health related updates, de-identified data request management, support for consistent healthcare management of patient groups or populations, and managing population health study related identifiers.

Record Infrastructure (RI)

This section consists of functionalities about record management. It can be implemented within the architecture of a single system or across tightly-coupled systems. There are 37 functionalities (including subfunctionalities) and 186 conformance criteria organized in 3 categories, including record lifecycle and lifespan, record synchronization and record archive and restore.

Trust Infrastructure (TI)

This section consists of functionalities related to EHRS trust infrastructure. These functionalities are fundamental to system operations, security, efficiency, data integration, privacy/confidentiality and interoperability with other systems. There are 93 functionalities and 681 conformance criteria organized in 9 categories, including security, audit, registry and directory services, standard terminology and terminology services, standards-based interoperability, business rules management, workflow management, database backup and recovery, and system and performance management.

Review of Chinese Health System Development

The Chinese government regard information technology as an important part of modern medicine. Health system construction is an essential part of health reform in China. The development of Chinese health information system can be divided into three stages, which are shown as below [6].

Stage 1 - Application of Computer Software

This stage is featured as the application of single computer software, such as finance management and medicine management. In this period, computer software is used to replace artificial operation. It is independent and the power of information technology is not sufficiently revealed.

Stage 2 - Construction of business systems

This stage is featured as the construction of various healthcare systems, such as hospital information system (HIS), community healthcare system, physical examination system, disease prevention and control system, and public health system. These systems realize the automatic process of information and increase work efficiency. Designed separately, they have simple data integration capability, and can only meet the needs of business activities.

Stage 3 - Construction of Regional Health Information Network

In this stage, the Chinese government began to develop regional health information network (RHIN) based on EHR [7]. The goal of RHIN construction is to break up the barriers of isolated systems, integrate existing healthcare systems, realize the interoperability and coordination, promote the exchanging and deep mining/analysis of EHR data. Currently, China will build 4 levels of RHINs, which are the national RHIN, provincial RHIN, municipal RHIN and county RHIN [8]. The RHINs are based on EHR and take the responsibilities of the center of health data exchange and application.

Weakness Analysis on Chinese Health System Construction

For a long time, Chinese health system construction has been driven by pure information technology and investment. It can only meet the needs of a specific domain. This construction mode lacks of top-level design and theoretical instructions, leading to a large number of isolated systems. To solve these problems, RHINs are built to create an infrastructure where health information can be shared, exchanged and analyzed. But a thorough integration is needed. That is, to apply HL7 EHRS FM to Chinese RHIN construction.

Suggestions on Applying HL7 EHRS FM in China

Broadly speaking, there are three suggestions on applying EHRS FM in China. They are shown as follows.

Enhancing the Functionalities of Chinese Healthcare Systems by Means of the EHRS FM

In the architecture of RHIN, the healthcare systems can only play a role of data provider. Their business functionalities are not enhanced by EHR. This leads to the functional stuffness of these systems. For example, a doctor cannot get the clinical history of patient when making a diagnosis. Therefore, it is needed to extend care provision, care provision support and administration support functionalities to hospital information systems and community healthcare systems. Meanwhile, it is needed to extend population health support functionalities to disease prevention and control/public health system. Table 1 gives the functionality list that can be applied to Chinese healthcare systems. It is noted that the existing functionalities of healthcare systems are not included.

Table 1 – Functionalities Applied to Chinese Healthcare Systems

Healthcare systems	Applied Functionalities
Hospital Information System	Management of demographics, clinical history, external-source data, workflow, business rules, patient privacy, correspondence, doctor, finance and administration
Disease Prevention and Control/Public Health System	Health risk notification, epidemiological investigation, donor management, health assessment, process improvement, healthcare management
Community Healthcare System	Future care management, Patient education support

Enhancing the Information Network Functionalities of RHIN

In China, the RHINs play a role of EHR data integration and exchange. But the systems to be connected are limited to HIS, disease prevention and control/public health system and community healthcare system. Only limited data types are supported by RHIN, such as structured data, documents and images. Compared with the EHRS FM R2, RHIN should extend the systems to be interfaced, such as physical examination systems or personal health record (PHR) systems. Secondly, the data types should also be extended to be compatible with medical instrument generated data. As an archive management center, the RHIN has EHR lifecycle management functionalities, such as creating archive, arranging archive, migrating archive, merge/split archive, final archive and archive maintenance. These archive functionalities should be extended as well. Table 2 shows the information network functionalities to be enhanced for RHIN in China.

Table 2 – Information network functionalities to be enhanced for RHIN in China

Functional Type	Functionalities to be Enhanced
Data Collection and Integration	Collect clinical documents, externally-sourced data, emergency medical system originated data, externally-sourced clinical images, patient-originated data
Record Synchronisation	Structured-document exchange, registry and directory services, complete logical record exchange
Record Lifecycle and Lifespan Management	Originate, manage, retain, amend, translate, verify, access, output, disclose, transmit, receive, de-identify/ re-identify, pseudonymize, extract, archive/restore, destroy/identify, deprecate/retract, re-activate, merge/unmerge, link/unlink record entries
Record Archive and Restore	Archive record entries from online data structures to near-line or off-line data structures, restore record entries

Redesigning the Trust Architecture of RHIN

The traditional concept of information security only contains security and audit. But the new trust infrastructure of HL7 EHRS FM R2 greatly extends this concept. It integrates system management functionalities, such as database backup and recovery and system and performance management. The registry and directory services, standard terminology and terminology services, standards-based interoperability, business rules management, and workflow management are also integrated into the new trust architecture. In China, security design is an important part of RHIN [9]. The security architecture of RHIN includes identity protection, identity authentication, identity management, access control, information encryption, digital signature, anonymization and security audit. Compared with HL7 EHRS FM R2, the security architecture of RHIN is not sufficient in registry and directory services, standard terminology and terminology services, and standards-based interoperability functionalities. Therefore, it needs to be redesigned according to the new trust infrastructure. The new trust architecture is shown in Table 3.

Table 3 – New Trust Infrastructure of RHIN in China

Functional Type	FunctionalityName
Security	Entity authentication, entity authorization, entity access control, patient access management, non-repudiation, secure data exchange, secure data routing, patient privacy and confidentiality, system operation measurements, service availability, trusted information exchange environment
Audit	Audit triggers, audit log management, audit notification and review
Registry and Directory Service	Registry services, directory services
Standard Terminology and Terminology Services	Standard terminology and terminology models, maintenance and versioning of standard terminologies, terminology mapping
Standards-Based Interoperability	Application, structured message, and structured document interchange standards, interchange standards versioning and maintenance, standards-based application integration, interchange agreements, system integration
Other Functionalities	Business rules management, workflow management, database backup and recovery, system management operations and performance

- [6] Rao Keqin. Electronic Health Record and Regional Health Information Network (Business Section). People Health Press, 2010:40-43
- [7] Rao Keqin. Electronic Health Record and Regional Health Information Network (Technology Section). People Health Press, 2010:504-536
- [8] Chinese National Health and Family Planning Commission. The 13th Five-Year Plan of National Population and Health Informationization: 2017.
- [9] Ni Ning. Security Analysis for Regional Health Information Network. Chinese Journal of Health Informatics and Management, **10**(2013):244-247

Address for correspondence

Dongsheng ZHAO
dszhao@bmi.ac.cn
(86)10-66931123

Conclusions

HL7 EHRS FM R2 is the latest advancement in EHRS functionality design. This version contains several new ideas. China should learn from this advanced international standard to improve the functionality design of health systems and promote the construction and development of RHINs in China.

Acknowledgements

Funding sources for the work and other relevant acknowledgements are noted here. Authors may also present disclosures or disclaimers to their work in this section.

References

- [1] International Standard Organization. Health Informatics - Electronic health record - Definition, scope and context. ISO/TR 20514: 2005.
- [2] ANSI/HL7. HL7 EHR-System Functional Model, Release 2: April 2014.
- [3] The Central Committee of the Communist Party of China, the State Council of China. Health China 2030: October 2016
- [4] International Standard Organization. Electronic Health Record-System Functional Model, Release 1.1. ISO/HL7 10781: 2009.
- [5] International Standard Organization. Electronic Health Record-System Functional Model, Release 2. ISO/HL7 10781: 2015.

Are Health Literacy and eHealth Literacy the Same or Different?

Helen Monkman^a, Andre W. Kushniruk^a, Jeff Barnett^a,
Elizabeth M. Borycki^a, Leigh E. Greiner^b, Debra Sheets^c

^a School of Health Information Science, University of Victoria, Victoria, British Columbia, Canada

^b BC Corrections, Victoria, British Columbia, Canada

^c School of Nursing, University of Victoria, Victoria, British Columbia, Canada

Abstract

Many researchers assume that there is a relationship between health literacy and eHealth literacy, yet it is not clear whether the literature supports this assumption. The purpose of this study was to determine if there was a relationship between health and eHealth literacy. To this end, participants' ($n = 36$) scores on the Newest Vital Sign (NVS, a health literacy measure) were correlated with the eHealth Literacy Scale (eHEALS, an eHealth literacy measure). This analysis revealed no relationship ($r = -.041$, $p = .81$) between the two variables. This finding suggests that eHealth Literacy and health literacy are dissimilar. Several possible explanations of the pattern of results are proposed. Currently, it does not seem prudent to use the eHEALS as the sole measure of eHealth literacy, but rather researchers should continue to complement it with a validated health literacy screening tool.

Keywords:

Health Literacy; Medical Informatics; Consumer Health Information

Introduction

Consumer health information systems and technology are increasingly popular. Citizens are actively seeking health information (e.g., diagnoses, medications, symptoms, treatment options) on the internet as well as using digital tools to help monitor and manage their health. In fact, nearly 6 in 10 (59%) Americans sought health information on the internet at least once during 2012 [1]. Further, consumers are increasingly accessing electronic personal health information (e.g., lab test results, prescriptions, medical history). For example, 2 in 10 Americans accessed their medical record at least once in 2014 [2]. However, this new medium of digital health communication, as well as increased access to personal health information brings challenges observed in traditional paper-based health information and create new challenges unique to using information technology.

It is concerning that an estimated 6 in 10 Canadians have limited health literacy [3]. Similarly, only one in ten Americans were considered to have proficient health literacy with the remainder having intermediate, basic, or below basic health literacy [4]. Health literacy is considered "the degree to which individuals can obtain, process, and understand the basic health information and services they need to make appropriate health decisions" [5]. Thus, people with limited health literacy may not have the necessary skills to acquire and understand health information as well as make optimal health-related decisions. In a review of the literature, limited health literacy was "consistently associated with increased hospitalizations, greater

emergency care use, lower use of mammography, lower receipt of influenza vaccine, poorer ability to demonstrate taking medications appropriately, poorer ability to interpret labels and health messages, and, among seniors, poorer overall health status and higher mortality" [6].

Given the potential implications of limited health literacy, the concept of health literacy has gained recognition as an important consideration for designing materials and interventions for health consumers.

Several different measures are available for measuring consumers health literacy skills. The scales vary in terms of their evaluation approach, administration time, and national healthcare context appropriateness. However, most health literacy scales generate an objective measure of a consumers' competency by assessing their skills (e.g., comprehension, numeracy, pronunciation of medical terminology). Currently, there is no consensus on the most appropriate health literacy measure [6].

Given the unique challenges associated with navigating healthcare in the digital era, the concept of eHealth literacy has gained traction and research attention. eHealth literacy is "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [7]. Norman and Skinner [7] proposed the Lily Model to depict eHealth literacy. This model is an amalgamation of six component literacies: computer literacy, information literacy, media literacy, traditional literacy and numeracy, scientific literacy, and health literacy [7].

Unlike health literacy, there is a dearth of tools available for assessing consumers' levels of eHealth literacy. Recognizing the value of aligning eHealth programs and the skills of their users and building on their Lily Model of eHealth literacy, Norman and Skinner developed the eHEALS [8]. Currently, this is the most commonly used tool for eHealth Literacy assessment [8]. The instrument "was designed for simple, easy administration and thus can be used on its own or incorporated with other measures of health as part of a standard health assessment battery in primary care or to support health promotion planning." [8]. Since its development, eHEALS has been used for various purposes and translated into other languages [e.g., 9, 10].

Developing measures for any construct can be challenging and health literacy and eHealth literacy are no exceptions. One challenge for measurement development is establishing its validity. Validity is defined as "the extent to which a measure reflects the concept. The measure reflects nothing more or less than that implied by the conceptual definition" [11]. Thus,

whether or not a tool actually measures what it intends to measure can be challenging.

Norman and Skinner identified the potential limitation of it as a self-report rather than empirical observation measure and the implication that it was an index of perceived rather than observed skills [8]. Given this potential shortcoming, the extent of the relationship between eHEALS scores and consumers' observed eHealth literacy skills has been investigated. One study found no evidence of a relationship between participants' ($n = 88$) objective performance eHealth tasks and their eHEALS scores [10]. That is, participants were presented explicit tasks and practical scenarios and then asked to solve a health problem such as "Why is the Swine flu not correct?" [10]. Given the dearth of eHealth literacy instruments and that the validity of eHEALS has been challenged, many researchers continue to rely on traditional health literacy tools for consumer health informatics research. Further, many researchers infer, either directly or indirectly, that eHealth literacy is grounded in health literacy, which in turn encompasses literacy. However, a discrepancy between literacy and health literacy was reported, whereby fewer people were identified as having low literacy than low health literacy, suggesting health literacy requires additional skills [3]. Thus, it is unwise to assume that health literacy is equivalent to eHealth literacy. Yet, one of the most popular health literacy assessments is based on whether participants are able to pronounce (i.e. a basic literacy skill) medical terms accurately [12] rather than on their comprehension (i.e. a health literacy skill). It is logical for researchers to assume there is a relationship between health literacy and eHealth literacy, given that (a) the primary difference in the definition of these two constructs is how health information is attained (i.e. paper-based vs. electronic resources) and (b) health literacy is a component of the Lily Model [7].

An assumed relationship between health literacy and eHealth literacy is also apparent in methodological approaches. Many researchers administer health literacy measures as a proxy for eHealth literacy rather than using measures developed to assess eHealth literacy itself. Often, this is unlikely to be problematic as these assessments are typically conducted in an applied consumer health information technology context. Thus, using health literacy as a proxy for eHealth literacy may be suboptimal, but other aspects of the study will still reveal whether or not participants have difficulty using electronic health information. Health literacy screening instruments have been reviewed for eHealth applications which is evidence of the popularity of this approach and supports this line of reasoning [13]. Therefore, the purpose of this study was to determine if there was a relationship between health literacy and eHealth literacy. It was hypothesized that there would be at least a moderate, positive correlation (i.e. $r > 0.4$) between participants' performance on the NVS (a health literacy measure) and eHEALS (an eHealth literacy measure).

Methods

Participants

This study recruited participants using faculty listserves and posters at the University of Victoria, British Columbia Canada. The participants ($n = 36$) ranged in age from 18 to 35 years old ($M = 23.6$ years $SD = 3.8$) and volunteered to participate. All participants completed both the NVS [14] and eHEALS [8] scales as part of a battery of tests in a larger study.

Apparatus and Materials

The NVS [14] is a 6-question scale used to measure health literacy through comprehension and numeracy skills. Participants must interpret information on a nutrition label to answer the questions correctly [14]. The range of possible scores on the NVS is zero to six and it requires approximately three minutes to administer [14]. Participants' scores are used to infer the likelihood of whether or not they are likely to have limited health literacy.

The eHEALS [8] is an 8-item self-report scale assessing consumers' impressions of their ability to seek, appraise, and apply health information gained from electronic resources [8]. The eHEALS asks participants how well they know health resources on the internet in terms of what is available, where and how to find helpful one, how to use them, how to evaluate them, how to differentiate between high and low quality ones, and how to use them to inform their health decision-making. Participants rate each question on a five point likert scale from Strongly Disagree (1 point) to Strongly Agree (5 points) and therefore the range of scores on the eHEALS is 8 to 40. However, Norman and Skinner [8] did not provide explicit instructions for differentiating between eHealth Literacy levels, merely that higher scores (i.e. more Agree or Strongly agree responses) indicate higher levels of eHealth literacy.

The demographic questionnaire and the scales were displayed on a 13 inch MacBook Air. Scales were entered into and administered using survey software.

Procedure

This study was a within-groups design and therefore all participants completed both scales. Participants were told to answer the questions to the best of their ability. If participants had questions, they were encouraged to ask the experimenter who remained nearby in the room. All participants completed the demographic questionnaire first, followed by the NVS [14], and finally the eHEALS [8]. All of the measures were administered online.

Descriptive statistics were computed for participants' demographics, NVS scores, and eHEALS scores. Additionally, a Pearson product-moment correlation coefficient was computed to assess the relationship between participants' NVS and eHEALS scores.

Results

Participants

Participants were predominantly female (72.2%, Table 1). As expected, all participants were students, (83.3%) most were enrolled full-time, and half (50%) of the participants reported high school as their highest level of education completed. Five participants (13.9%) spoke a first language other than English.

Table 1 – Sample Characteristics

Descriptor	Categories	N (%)
Gender	Female	26 (72.2)
	Male	10 (27.8)
Highest Completed	High School	18 (50.0)
Education Level	Undergraduate	13 (36.1)
	Graduate	4 (11.1)
	Other	1 (2.8)
First Language	English	31 (86.1)
	Other	5 (13.9)

Newest Vital Sign (NVS) Scores

Participants scored an average of 5.1 (Range 2 to 6, SD = 1.2) on the NVS [14]. The NVS had poor internal consistency ($\alpha = 0.57$). However, this is likely the result of the limited number of items in the scale as well as the dichotomous (i.e. correct or incorrect) nature of the questions.

The majority of the participants ($n = 30$, 83%) scored five or six on the NVS, indicating a very low likelihood of having limited health literacy (see Figure 1) and further suggesting that there was a ceiling effect. Six (17%) of the participants had NVS scores of 3 or fewer, indicating the possibility they had limited health literacy. However, there was no indication any participants had a high likelihood of limited health literacy, as no participants scored 0 or 1.

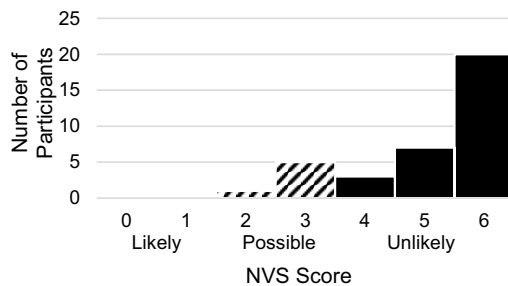


Figure 1 – Participants' NVS Scores and the Implied Likelihood of Limited Health Literacy

eHEALS Scores

Participants scored an average of 27.9 (Range 17 to 38, SD = 5.5) on the eHEALS, which demonstrated good internal consistency ($\alpha = 0.86$).

Table 2 – eHEALS Responses (Scores)

Question / Statement	Mean	SD
* How useful do you feel the Internet is in helping you in making decisions about your health?	3.8	0.93
* How important is it for you to be able to access health resources on the Internet?	3.9	0.89
1. I know what health resources are available on the Internet	3.5	0.94
2. I know where to find helpful health resources on the Internet	3.3	0.93
3. I know how to find helpful health resources on the Internet	3.7	0.79
4. I know how to use the Internet to answer my questions about health	3.7	0.98
5. I know how to use the health information I find on the Internet to help me	3.3	0.97
6. I have the skills I need to evaluate the health resources I find on the Internet	3.4	1.08
7. I can tell high quality health resources from low quality health resources on the Internet	3.7	0.78
8. I feel confident in using information from the Internet to make health decisions	3.3	0.95

* supplementary questions, not formally part of the eHEALS

A framework was developed to make inferences about participants' scores on the eHEALS. The neutral score was deemed to be the value a participant would earn by answering all eight eHEALS questions with a neutral response (i.e. 24). Two categories were created on either side of this neutral value to postulate the confidence participants had in their eHealth skills based on their eHEALS scores (see Table 2). A histogram was constructed to assess the distribution of participants' eHEALS scores based on the categories generated (see Figure 2).

Table 3 – eHEALS Score Categories

Level of Perceived eHealth Literacy Skills	Responses	Score Range
Lack	Predominantly Disagree	8 – 15.99
Low	Mostly Disagree	16 – 23.99
Moderate	Mostly Agree	24 – 31.99
High	Predominantly Agree	32 – 40

Interestingly, 7 participants (see striped column in Figure 2) predominantly disagreed with eHEALS statements, which suggests they were doubtful of their eHealth literacy skills. The majority of the sample (21, 58.3%) perceived themselves as moderately capable of performing eHealth tasks (Table 3). Only 8 participants scored between 32 and 40 indicating high confidence in their eHealth capabilities.

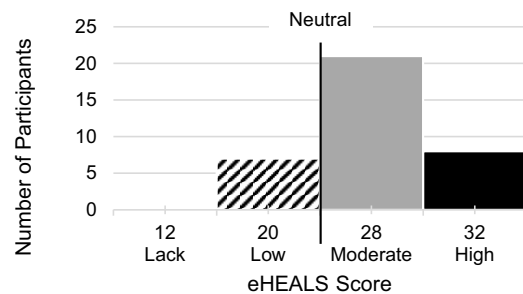


Figure 2 – Participants' eHEALS Scores

Relationship Between NVS and eHEALS Scores

A Pearson product-moment correlation coefficient was computed to assess the relationship between participants' NVS and eHEALS scores. No significant relationship ($r = -0.041$, $p = 0.81$) was observed between the two variables (see Figure 3). This suggests that the relationship between eHEALS and NVS is limited or non-existent.

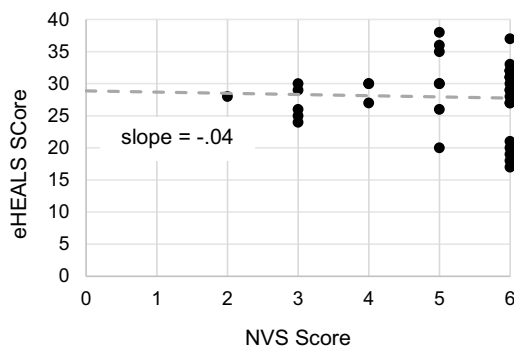


Figure 3 – Scatterplot of Participants' eHEALS and NVS Scores

Interestingly, six out of seven participants identified as having low confidence in their eHealth Literacy skills scored high (5 or 6) on the NVS. This pattern suggests that despite having strong observed health literacy scores, these participants were doubtful of their ability to accomplish eHealth tasks.

Discussion

Contrary to the hypothesis, no relationship between participants' scores on the NVS and eHEALS was observed.

Moreover, participants had generally high levels health literacy scores, but their eHealth literacy scores were much more variable. There are several different explanations for this pattern of results.

This finding could be the result of health literacy and eHealth literacy being very distinct concepts. That is, contrary to the hypothesis that health literacy is the foundation of eHealth literacy, the skills required for engaging with electronic health resources are so unique from those required for interacting with paper materials, there is no relationship between an individual's health literacy and eHealth literacy.

Another possible explanation is that the eHEALS assesses health information seeking and appraisal, whereas the NVS only requires comprehension and application of health information. That is, the NVS provides participants with the health information necessary to answer the questions posed, rather than having to find it and determine its trustworthiness. However, the eHEALS asks participants about all three aspects with respect to health resources on the internet.

These results may have been observed because the eHEALS is a self-report measure, which is a suboptimal measure of actual skills. This argument is supported by evidence that participants' scores on the eHEALS were not positively related to their performance on actual eHealth tasks [10]. In most circumstances, the weaknesses of self-report measures are that there is a tendency for respondents rate themselves in a more socially desirable (or favourable) manner than what is true in reality. However, the pattern of results from this study suggest that participants are actually more critical of their eHealth literacy skills than one would expect them to perform. That is, despite being able to apply health information successfully (i.e. high NVS scores), participants may be overly doubtful of their eHealth skills. Participants in this study were all university students and as such they may approach information on the internet more cautiously.

The final explanation is that eHEALS is not a valid measure of eHealth literacy. The validity of this tool was previously challenged by the lack of relationship between eHEALS and performance on eHealth tasks [10]. Logically, eHealth task performance and eHEALS scores would have a strong positive correlation if eHEALS actually measured eHealth literacy. In contrast, because health literacy is merely one of eHealth literacy's multiple facets, a moderate correlation would be expected. However, neither of these expected patterns of results was observed. The weakness of eHEALS may simply be that it is a self-report measure. That is, people may be substantially more or less confident (suggested by this data) in their ability to seek and appraise eHealth resources as well as apply the information contained therein. However, even if eHEALS measures consumers' confidence using eHealth tools rather than their actual competency, it still has value. For example, people who are less confident may require more support with electronic health interventions even though they are capable of using them successfully.

Norman addressed some of the identified shortcomings and inconsistencies in research published since his work with Skinner (i.e. Lily model, eHEALS). Specifically, he argued that the concept of eHealth literacy warranted reconsideration in the emerging context of Web 2.0 and popularity of social media and that this dynamic eHealth landscape explains why eHEALS scores were not correlated with eHealth tasks [15]. Further, Norman asserted that eHealth literacy should be considered a "form of meta-literacy, combining many different literacy skills beyond just health literacy or numeracy". Norman cautioned against extrapolating results from one or two component literacies to make claims about eHealth literacy (as done in this study), because eHealth literacy should be considered a set of

integrated rather than independent skills [15]. Despite this warning, scores on health literacy measures are often used as a proxy for eHealth literacy skills in studies of consumer health informatics. It is unfair to blame the researchers, because if the eHEALS is not a valid measure of the integrated eHealth skills, they are forced to adopt an alternative. Further, as previously argued, given the similarity of the definitions for health and eHealth literacy, it is sensible to assume that there should be substantial overlap between these two constructs.

The sample for this study was predominantly female, early adult, full-time university students, who spoke english as their first language. Given their age and education level, it is fair to assume this sample is technologically skillful. Further, with this fairly homogeneous sample, one would expect high levels of both health and eHealth literacy. This group demonstrated health literacy proficiency with the majority earning high scores on the NVS (i.e. a positively skewed distribution) and evidence of a ceiling effects. In contrast, participants' eHEALS scores yielded an interesting pattern of results. Specifically, it was expected that the majority of eHealth Literacy scores would indicate the sample had high confidence in their skills. Higher eHEALS scores (i.e., more agree and strongly agree responses) suggest high confidence in participants' ability to find, evaluate, and use health resources on the Internet. However, eHEALS scores from this sample of participants were much more variable than expected and more closely resembled a normal distribution as opposed to being positively skewed.

Responses to the supplementary questions suggest that participants feel that the Internet is a useful resource for making health decisions and it was important for them to be able to access health information resources on the Internet. However, the three lowest rated formal eHEALS questions (see Table 2) suggest participants were not as confident in:

- Where to find helpful health resources on the Internet?
- How to use health information on the Internet to help them?
- Using information from the Internet to make health decisions.

Based on their total eHEALS scores, most participants had moderate confidence in their eHealth skills. Almost equal parts of the remaining participants had either high or low confidence in their eHealth skills.

One limitation is that only the paper-based NVS has been validated. However, it was administered digitally in this study. Additionally, the Chronbach's alpha was low indicating poor internal consistency of the NVS. Another limitation was the relatively small size.

Conclusion

Results from this study indicate that people one would expect to be very confident in their eHealth literacy skills (i.e. young, well educated), many people are only moderately confident and some even doubtful. One of the motivations for assessing eHealth literacy is to assess how suitable an eHealth interventions might be for a health consumer. If eHealth literacy skills are low in this demographic, it is likely much worse in populations who more likely to have limited health literacy (e.g., seniors, immigrants). This finding also prompts us to question whether the concepts of health literacy and eHealth literacy need to be revisited, new measures needs to be developed, or both.

If the eHEALS is not a valid measure of eHealth literacy skills, it should not be used as the sole index for eHealth literacy. However, it could still be valuable to use an assessment of

consumers' confidence with eHealth tasks. As a precautionary measure, it would be wise for researchers to continue to administering health literacy screening tools paired with eHEALS in consumer health informatics research. Typically, administering a health literacy measure and eHEALS is not time prohibitive. However, this approach does not serve as a solution.

Despite development of a new eHealth literacy measure [16], and another in development based on a more robust eHealth literacy framework [17], these measures likely have the same weakness as the eHEALS: consumers' may underestimate or overestimate their eHealth skills, limiting the validity of self-report measures. Researchers are in need of a valid, objective (rather than self-report), rapid measure of eHealth literacy. Perhaps this could be done in by standardizing tasks similar to those in Van der Vaart's [10] study. However, this type of measure would have been developed with careful consideration to control or standardize the test (i.e. participants experience the same events with every administration). The Internet is dynamic and thus, participants attempting tasks in a live environment would introduce variability (e.g., different search results). However, it is important that the tasks used as indices of eHealth literacy are realistic. A simulated internet environment could be used as experimental control over the experience of each participant. Given that health inquiries on the internet predominantly begin with a search engine [1], this is an important eHealth literacy task. Including a search task and requiring participants to assess the quality of search results would be a useful and realistic eHealth task. The skill set required for health consumers to navigate and benefit from an increasingly digital landscape is undoubtedly unique from a paper based world. However, establishing a valid measure for eHealth literacy remains an outstanding challenge.

Currently, due to the different measurement approach of the tools (i.e. objectively vs. self-report), it is difficult to discern whether there truly is or is not a relationship between health literacy and eHealth literacy. Future research is required to investigate whether health literacy skills are fundamental to, have some overlap with, or are entirely different from eHealth literacy skills.

References

- [1] S. Fox S and M. Duggan, Health online 2013, Pew Research Center's Internet & American Life Project (January 15, 2013): Washington DC
- [2] V. Patel, W. Barker, and E. Siminerio, Trends in consumer access and use of electronic health. ONC Data Brief, no. 30, (October 2015), Office of the National Coordinator for Health Information Technology: Washington DC
- [3] Health literacy in Canada: A healthy understanding, (2008), Canadian Council on Learning: Ottawa ON
- [4] M. Kutner, E. Greenberg, Y. Jin, and C. Paulsen, The health literacy of America's adults: Results from the 2003 national assessment of adult literacy (NCES 2006-483), (2006), U.S. Department of Education. Washington, DC: National Center for Education Statistics.
- [5] U.S. Department of Health and Human Services. 2000. Healthy People 2010. Washington, DC: U.S. Government Printing Office. Originally developed for S.C. Ratzan, and R. M. Parker. 2000. Introduction. In National Library of Medicine Current Bibliographies in Medicine: Health Literacy. C.R. Selden, M. Zorn, S.C. Ratzan, R.M. Parker, Editors. NLM Pub. No. CBM 2000-1, (2000), Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services.
- [6] N.D. Berkman, S.L. Sheridan, K.E. Donahue, D.J. Halpern, A. Viera, K. Crotty, A. Holland, M. Brasure, K.N. Lohr, E. Harden, E. Tant, I. Wallace, M. Viswanathan, Health literacy interventions and outcomes: An updated systematic review, (March 2011), Evidence Report/Technology Assessment No. 199. (Prepared by RTI International—University of North Carolina Evidence based Practice Center under contract No. 290-2007-10056-1. AHRQ Publication Number 11- E006. Agency for Healthcare Research and Quality: Rockville, MD.
- [7] C.D. Norman, and H.A. Skinner, eHealth literacy: essential skills for consumer health in a networked world, *J Med Internet Res* 8(2), (2006), e9.
- [8] C. Norman, and H. Skinner, eHEALS: the e-Health Literacy Scale, *J Med Internet Res*, 8(4), (2006), e27.
- [9] Koo, M., Norman, C. D., & Hsiao-Mei, C. (2012). Psychometric evaluation of a chinese version of the eHealth Literacy Scale (eHEALS) in school age children. *Global Journal of Health Education and Promotion*, 15(1).
- [10] R. Van der Vaart, A.J. van Deursen, C.H., Drossaert, E. Taal, J.A. van Dijk, and M.A. van de Laar, Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations, *J Med Internet Res*, 13(4), (2011), e86.
- [11] W. Jackson, and N. Verberg, Methods: Doing social research (4th ed.), (2007), Toronto: Pearson Prentice Hall.
- [12] Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E. J., George, R. B., Murphy, P. W., & Crouch, M. A. (1993). Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Family medicine*, 25(6), 391-395.
- [13] Collins, S. A., Currie, L. M., Bakken, S., Vawdrey, D. K., & Stone, P. W. (2012). Health literacy screening instruments for eHealth applications: a systematic review. *Journal of biomedical informatics*, 45(3), 598-607.
- [14] B.D. Weiss, M.Z. Mays, W. Martz, K.M. Castro, D.A. DeWalt, M.P. Pignone, ... and F.A. Hale, Quick assessment of literacy in primary care: the newest vital sign, *The Annals of Family Medicine* 3(6), (2005), 514-522.
- [15] C. Norman, eHealth literacy 2.0: problems and opportunities with an evolving concept, *J Med Internet Res* 13(4), (2011), e125.
- [16] G. Seçkin, D. Yeatts, S. Hughes, C. Hudson, and V. Bell, V. Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument. *J Med Internet Res*, 18(7), (2016), e161.
- [17] O. Norgaard, D. Furstrand, L. Klokke, A. Karnoe, R. Batterham, L. Kayser, and R. Osborne, The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems, *Know Man E-Learn*, 7(4), (2015), 522-540.

Address for correspondence

Helen Monkman,
Assistant Teaching Professor
School of Health Information Science
University of Victoria, Victoria, BC, Canada,
email: monkman@uvic.ca

An Exploratory Analysis of Game Telemetry from a Pediatric mHealth Intervention

Rema Padman^a, Disha Gupta^a, Bhargav Sri Prakash^b,
Chelladurai Krishnan^b, K. Panchatcharam^b

^a The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA,

^b FriendsLearn, Chennai, India

Abstract

Pediatric obesity is a growing epidemic, with unhealthy eating habits and poor physical activity being major contributors. While video and mobile games have been shown to have a positive impact on behavior change in children, the mechanisms underlying game play that impact outcomes of interest are poorly understood. This research aims to examine the impact of a novel mobile gaming app on the design of behavioral interventions by learning from the rich and unique game telemetry generated from a randomized controlled trial of the app use by school children. In this exploratory analysis, we extract a partial dataset to build and analyze chronological sequences of game plays to understand key patterns in the game mechanics that players utilize as they navigate the game, and possible implications of the results.

Keywords:

Pediatric Obesity; Mobile Apps; Game Telemetry Analysis

Introduction

This research is motivated by the need to explore some of the challenges and related opportunities at the intersection of the rising epidemic of pediatric obesity worldwide [1], healthy eating and physical activity related behavior modification challenges in children [2], and the growing role of gamification and learnification on mobile devices in the lives of these pediatric digital natives [3]. While there is some early evidence of the positive impact of video games on healthy eating behaviors in children [4], the mechanisms underlying these improved outcomes are yet to be understood. To design appropriate interventions in the game environment for children's behavior formation and change, we need to learn more about the underlying patterns of player behaviors evidenced during gameplay, the goals they are trying to achieve with the game, and their overall gameplay experiences.

In this exploratory study, we analyze game telemetry to understand user interactions from playing Fooya, an iOS/Android based mobile App that has been shown to improve the nutrition-health of children through virtual reality-based immersive mobile gaming which uses Artificial Intelligence to achieve personalized behavior reinforcement [5]. We analyze the interactions with the game choices made by the players in a randomized controlled experiment in a middle school setting. Learning from and associating game playing behaviors with known factors affecting unhealthy eating behaviors and the players' food choices at the end of the game may provide new insights into the complex interactions between game playing and health behavior changes necessary to design more impactful interventions via games on mobile devices in order to improve pediatric overweight and obesity rates.

Background

Pediatric Obesity

Obesity is an increasingly common epidemic in children. The growing pediatric obesity rate worldwide has serious health consequences such as cancers, diabetes, asthma and shortened life span as well as higher healthcare costs [1; 2; 6]. While there are many factors that contribute to overweight and obesity, diet decisions are a leading cause [1]. Unhealthy eating habits and poor physical activity is a major contributor to the growing obesity epidemic [7]. There is significant ongoing focus on addressing the risk factors and moving the population to a healthier lifestyle through health education/ communication and motivation. Establishing healthy eating habits early in life is important because childhood habits are predictive of those in adulthood [7]. There is a clear need to identify effective methods for improving dietary intake and physical activity habits earlier in life that are acceptable to children.

Games on Mobile Devices

Games have been shown to have positive impact on children [8]. Many children spend several hours a day playing video games [3; 4], mostly on mobile devices, a platform through which they can learn about health in a fun and enjoyable way. This may be one approach to address the multi-faceted challenge of the pediatric overweight and obesity epidemic [9, 10].

However, a critical barrier to progress in the field is the lack of understanding about the mechanisms underlying games that impact outcomes of interest. Video games include many levels of challenges, imaginative virtual worlds, and the opportunity to navigate them in distinct ways, alone or in teams [11]. The appeal of discovery may be a strong motivator of these games to some players while others may enjoy the competitive or collaborative nature of the game. Gamification is the studied, thoughtful, and creative application of game design elements to engage the player [11; 12]. Games with a highly directed experience and tiered set of tasks can challenge and motivate players. While much is known about game design, recent research has highlighted the gap in understanding the specific mechanisms that link the game playing behaviors with observed outcomes so that game design as well as re-design can be informed through evidence based knowledge and practices [12-14].

Game Telemetry

Large volumes of data are routinely collected during game play and analyzed in simple ways. Game developers nowadays remotely and unobtrusively monitor every aspect of a game, allowing them to accumulate large amounts of data of the player-game interaction over extended time periods [15]. One type of data collected is the actual clicks made by the player as

they navigate the game, called game telemetry. At the current time, this clickstream data is used to analyze questions such as whether players are entering the game or not, if they are dropping out early or late, if they are playing the game for a while and then dropping out at some distinct step, and so on. This type of feedback is made available in real time so that game developers can track usage and player engagement that have implications for success of the game.

Learning from clickstream/telemetry data.

This data is a valuable source for developers and game designers to guide decision-making throughout the game design process, to understand player movements, reduce production costs, or uncover bugs, among others. At the same time, the increasing popularity of online gaming has led to in-game statistics to improve player experiences by providing performance summaries and comparisons to others playing the game. There is now a growing field of game analytics, some of it focused on using telemetry data to discover and communicate meaningful patterns in data in the context of game development and game impact [15]. For example, identifying which paths result in the longest game time and what game features engage players in these paths require methods to learn paths from the detailed telemetry data and then examine the features that define these paths in unique ways. Hence new approaches that leverage advanced statistical machine learning methods that can learn from large volumes of multiple streams of high dimensional data have the potential to facilitate the analysis of patterns in telemetry data streams.

Impact on design of behavioral interventions.

There is a growing recognition of the need for more refined strategies for designing behavioral interventions for complex health conditions that allow incorporation of personalized evidence-based knowledge and increased use of data and information technology to improve health outcomes [16]. Researchers have suggested that behavioral science can provide new insights to make key design decisions by the designers of serious video game for health-related games. Future research needs to investigate the most effective ways to achieve the dual goals of fun-ness and seriousness [14]. The availability of highly granular game telemetry from mobile video games promoting healthy behavior changes, enhanced with data about food habits and choices, provides a unique opportunity to develop data-driven, statistically rigorous, models and methods for learning from this data to provide insights for designing interventions in mobile video games for diverse health conditions that can be further evaluated in actual decision settings [17]. For example, real-time game data can inform designers where players are having difficulties in the game and can potentially modify the design, perhaps in real time, to improve that experience.

Methods

Mobile gaming App

fooya!™ is a novel mobile gaming, iOS/Android based App that has been shown to improve the nutrition-health of children through virtual reality based immersive mobile gaming that uses Artificial Intelligence to achieve personalized behavior reinforcement [5]. fooya!™ has been shown to deliver statistically significant outcomes, with respect to food choices,

during randomized-controlled clinical trials conducted at the Baylor College of Medicine's Children's Nutrition Research Center [5]. Researchers at Kaiser Permanente Medical Group were involved in analyzing longitudinal impact on over 1000 children at multiple sites over a weeklong period during the ExxonMobil Summer Science Camp [18]. These findings are significant as increasing awareness and self-efficacy are the first steps to achieving health behavior change [19]. Outcomes from these and other trials have shown consistently positive trends, while also exposing opportunities for product feature development, enhancement and continuous refinement. This novel approach to behavior design aims to pave the way for targeting lifestyle disease prevention by enabling lifelong healthy lifestyle habits. This may be achieved by instilling healthier preferences during early childhood and the formative years of life.

Through game design that achieves a psychological state of flow for deep cognitive engagement, fooya!™ aims to increase awareness and induce greater self-efficacy. Underlying health models enable story telling surrounding topics of metabolic equilibrium and health balance pertaining to the user's Avatar. fooya!™ harnesses multiple core game mechanisms that deliver a single player and multiplayer experience to achieve behavior design through a social, mixed reality experience. Game compulsion loops guide users while getting harder over time, which makes users think more deeply and strategically, the more they play. The core engagement mechanism in fooya!™ adaptively engineers user experience moments that deliver a highly contextual message through a method of experiential discovery, as the result of learnification – a design thinking method of embedding and harnessing contextual learning within core entertainment. Based on a hypothesis derived from pediatric neuropsychology, fooya!™ was developed to deliver innovative therapeutic entertainment to make healthy behavior change incredibly fun for the children.

While the experimental studies with fooya!™ have shown promising results as a result of innovations through a Digital Vaccine candidate technology based on Neuropsychology to reduce the risk of lifestyle diseases [20], they fail to explain the mechanisms underlying these improved outcomes. To design appropriate interventions for behavior formation and change via video games, we need to learn more about the underlying mechanisms that link game playing to healthy/unhealthy eating behaviors. Attracting and maintaining a child's attention may be the biggest contribution of video games to health-related behavior change, but this has not been demonstrated [13]. Game playing is a complex process. The automation of analysis and visualization of game telemetry is an opportunity for Artificial Intelligence based personalization for specific users.

Data

A rich and unique data set was generated via a randomized controlled trial of the use of fooya!™ by school children. With ethics board approval and informed consent from participants, a research study was conducted to assess the awareness levels among urban Indian children regarding diet and lifestyle behaviors and evaluate the influence of the fooya!™ digital vaccine intervention, among school-age children in India. Specifically, the study objectives were to quantify the effectiveness of fooya!™ on health awareness around eating right and physical activity; find out the current diet and physical activity among urban children in India and the factors that affect them; and assess the extent of their awareness about eating right and physical activity. Using a pre-test, post-test research design, a total sample of 90 students from 3 urban schools

between 10-11 years of age, equally split between control and treatment groups and randomly assigned, participated in the study. A structured questionnaire first collected demographics, food habits, self-efficacy in selecting healthy foods, nutritional knowledge, and use and frequency playing video games using validated measures from the literature. Students in the treatment group played the mobile game for up to 20 minutes, while the control group played a board game. Once the game playing was completed, the participants were offered a choice of snacks to assess behavioral health outcomes, and a post-test questionnaire was administered to collect information similar to the pre-test questionnaire. All telemetry data associated with the actual game play by the children was also gathered and de-identified for the purposes of the study.

In the game, children make several decisions with split second timing, such as food choices, destroying bad/unhealthy food robots using the bad foods that are thrown at the player, and saving themselves. If the children collect good/healthy foods, they are in *fit-zone* for a while, which shields them from bad food robots. Throughout the game, telemetry data is collected across different game levels about bad foods thrown by robots at the player, good foods collected by the player during the game, how many times children read 'nutrition facts' and of what kind of food, for how long they were in *fit-zone*, and how many times they were hit by bad food robots. The activities in each level also sequentially influence activities in the next level, potentially affecting children's behavior enough to influence their real food choices eventually. By finding patterns in this game playing data, we can investigate the dominant and rare navigation patterns and their paths to the food choices of children. Eventually, a feedback loop in the game design may be needed to also influence them at the right time to make better and healthier food choices. Figure 1 shows a screen shot of the game with all the mechanics detailed above.



Figure 1 – A screen shot of fooya!™

Our exploratory analysis utilized 2 primary sources of data:

(1) Gameplay measurement variables from clickstream data that included:

- Total gameplay time, measuring game engagement: Per level and Overall
- Total levels per session, measuring performance outcomes
- The extent of customization or difference from the stock character will be assessed to indicate the extent of vicarious experience/self-modeling. The characteristics considered include: character gender, character clothing (female), character clothing (male), and so on.
- Average movement per minute, measuring game engagement; Average time: Per level and Overall
- Setting type (by level) for contextual analysis

- Type of nourishing food “consumed” (e.g., broccoli, cauliflower)
- Frequency of type of nourishing food “consumed” (e.g., broccoli, cauliflower)
 - Hit to Miss ratio (Accuracy of throws) measuring performance outcomes: Per level and Overall
 - Number of times hit, measuring performance outcomes: Per level and Overall
- Net calories accumulated per level, measuring performance outcomes
- Type of ammo obtained, measuring vicarious experience
- Active selection of healthy power-ups, measuring vicarious experience, including type and frequency: Per session and level
- Type of robots encountered, measuring performance outcomes, including number met and ratio of number destroyed and number met

(2) The second source of data is the demographic data on the players, meta data on the game plays such as playing time, completion status, and so on, and the food choices of the children at the end of the game.

In this exploratory analysis, we extracted a small subset of the de-identified game play data of 14 children who participated in the study in July-August 2016. We conducted a simple descriptive analysis of game mechanics to better understand the flow and critical features of the game in order to design advanced analysis of the play sequences of the full cohort in a future study. In particular, we compare the game sequences of the 14 players in their first level with that in the final level and summarize these differences, though the small sample size limits generalizability of the results.

Analysis

Construction of Event Sequences

Each player's telemetry data, described earlier, was processed and encoded according to a feature labeling scheme by which every possible feature currently utilized in the game by the players was assigned a distinct feature label. These labeled events are ordered chronologically based on their recorded timestamps. These chronologically ordered sequences of events represent the game playing experience of each player. Both game mechanics, which are techniques refined by designers to engage users in gameplay such as intangible rewards and recognition for achievements, and game dynamics, which include techniques designed to affect the pace of gameplay, such as time limits or countdowns, are used in fooya!™.

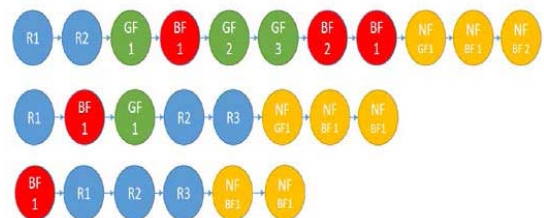


Figure 2 – Example Game Play Sequences of 3 Children

Figure 2 displays the game play sequences of 3 children, illustrating both game mechanics and game dynamics. Each path represents multiple levels in the game, with each level comprising different types of interactions. The blue boxes represent robots that are destroyed by the player, the green boxes represent the good/healthy foods they grab during the game and the red boxes represent bad/unhealthy foods that attack the player which they need to dodge. Lastly, the yellow boxes represent the nutritional facts read by the child. Across different players and their levels of play, the number, type and sequences of nodes can greatly vary and is reflective of their game engagement.

For example, in sequence 1, the player begins the game by destroying 2 robots, then grabs a healthy food item, is attacked by a bad/unhealthy food item, grabs 2 good food items, is attacked again 2 bad food items and ends the level by reading nutrition facts about three different foods, one of healthy type and 2 of unhealthy type. Sequence 2 of the second player includes fewer actions within the same time limit, while sequence 3 of player 3 has the least and with no access to good food items or nutrition facts on good foods.

In this preliminary study, we first conduct a descriptive summary of the game mechanics and dynamics that are demonstrated by the sequences. Together, this analysis will provide some early insights into the patterns that are utilized by the players as they navigate through the game's mechanics, reflecting game playing processes and its flow.

Results

The 14 players were equally divided in terms of gender, and the ages of the players were very similar since they were all recruited from the same grade at school. A descriptive summary of the game telemetry displayed in Figure 3 shows that some children played only 3 levels of the game, while others went much farther, with 1 child playing 12 levels of the game over the same duration as the others, though on average, males and females played similar number of levels. Figure 3 also summarizes the game mechanics used by the players in the first and final levels of the game. The aim was to examine changes in the access to different mechanics between these levels by all players, and the differences, if any, by gender. We observe that there is a significant increase in the use of different mechanics between the first and final levels, indicating that the players may be learning to play the game better as they navigate the levels. Furthermore, the differences between male players and female players also increase, particularly in the case of bad foods and robots destroyed.

		Average			
		Lower bound	Upper bound	Total	Average
First Level	Number of levels	3	12	8.5	8.9
	Nutritional Facts	1	11	3.6	3.6
	Good Foods	0	4	1.6	2.3
	Bad Foods	0	14	2.6	1.6
Last Level	Robots Killed	0	14	5.2	8.3
	Nutritional Facts	0	6	1.9	1.9
	Good Foods	0	6	2.3	3.3
	Bad Foods	1	47	12.7	18.9
		Lower bound	Upper bound	Total	Average
		1	23	11.1	17.0
					5.3

Figure 3 – Summary of game mechanics in the first and final levels of game play across all 14 players

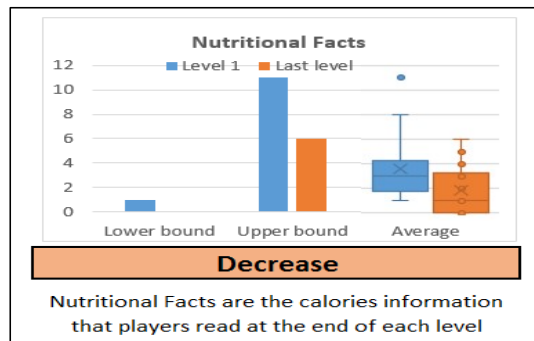


Figure 4 – Change in average number of Nutritional Facts looked up by the players in first and final levels of their games

Figures 4, 5, 6 and 7 display the changes in each of the game features of fooya!™ between the initial and final levels played. Interestingly, fewer Nutritional Facts are looked up by the players, as shown in Figure 4. This may likely be due to the increasing familiarity that the players feel regarding the nutritional content of the foods they encounter which decreases their interest and motivation in looking up this information over and over again, but this needs to be verified across other levels of the game and the post-intervention survey responses. This can be useful as early feedback for game re-design that may require new features to capture and retain the interest of the players in a critical component of the game's mechanics, which is about teaching nutritional facts associated with good and bad foods in a fun way.

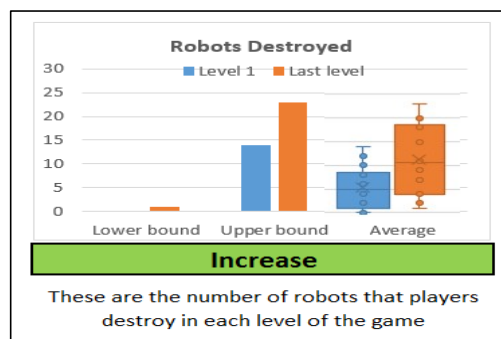


Figure 5 – Change in average number of Robots destroyed by the players in the first and final levels of their games

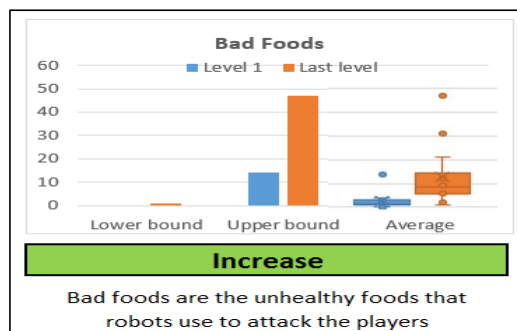


Figure 6 – Change in average number of unhealthy foods used to attack players in the first and final levels of their games

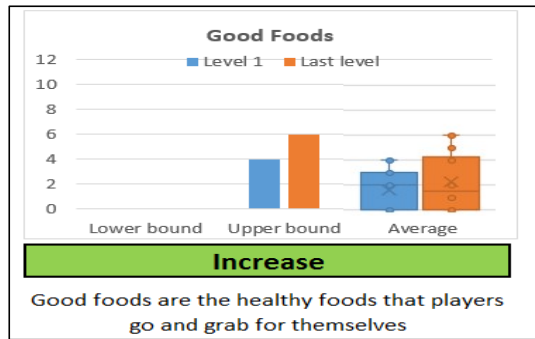


Figure 7 – Change in average number of healthy foods that players grabbed in the first and final levels of their games

Figure 5 shows almost a doubling in the average number of robots destroyed by the players between first and final levels of their games, indicating that the children are learning to identify the bad foods thrown at them and respond to it quickly to save themselves. Consistent with this implication, Figure 6 shows that there is also an increase in the average number of bad foods utilized during the games, so the players are actively engaging with the game mechanics as they proceed from first level to the final level. Finally, Figure 7 displays an increase in the average number of good foods selected by the players, likely implying that there may be some learning occurring regarding healthy vs. unhealthy foods.

Discussion

This exploratory study examines some preliminary game sequences of children playing a novel healthy eating related mobile app to gain a better understanding of the use of game mechanics. However, how this may be correlated with food choices at the end of the game is yet to be investigated. The current analysis provides some early indications of players' interests in using the different game mechanics built into the game and the need to design additional methods to display nutritional information that continues to engage the players over all levels of the game. Due to the limited number of participants analyzed in this study, it is not possible to generalize the results at this time.

Conclusion

Analysis of game play is challenging due to the multiplicity of game features, mechanics, dynamics and other characteristics. Future research will investigate the use of additional data from a larger set of participants to explore all aspects of game play and their associations with decisions made by the players regarding their food choices and the likely reasons for them based on survey responses. The potential impact of these decisions on pediatric obesity and health costs is a longer term investigation of interest.

Acknowledgements

We are grateful to Dr. Pradeep Krishnatray and Uttara Bharath Kumar from the Bloomberg School of Public Health at Johns Hopkins University, along with their team of researchers at the Center for Communication & Change – India, who conducted the randomized trial of the app that generated the clickstream data used in this study.

References

- [1] Ogden CL, Carroll MD, Kit BK, Flegal KM. (2014). Prevalence of childhood and adult obesity in the United States, 2011-2012. *Jama*, 311(8), 806-814.
- [2] Levi J, Segal LM, St Laurent R, Lang A, Rayburn JF. (2014). *F as in Fat: How Obesity Threatens America's Future*, 2012. Washington, DC, and Princeton, NJ: Trust for America's Health and Robert Wood Johnson Foundation.
- [3] Roberts D, Foehr U, Rideout V. Generation M: media in the lives of 8-18 year olds. Kaiser Family Foundation Study, 2005:12.
- [4] Charlier N, Zupancic N, Fieuws S, Denhaerynck K, Zaman B, Moons P. Serious games for improving knowledge and self-management in young people with chronic conditions: a systematic review and meta-analysis. *J Am Med Inform Assoc* 2016; 23:230–239. doi:10.1093/jamia/ocv100.
- [5] Johnston, C., & Moreno, J. (2014). The impact of a nutrition education mobile game on snack selection of middle school students. <http://obesity-week.com/wp/uploads/2014/10/TOS-Posters.pdf>
- [6] Wright DR, Prosser LA. The Impact of Overweight and Obesity on Pediatric Medical Expenditures. *Appl Health Econ Health Policy* 2014; 12:139–150
- [7] Lytle LA, Seifert S, Greenstein J, McGovern P. How do children's eating patterns and food choices change over time? Results from a cohort study. *Am J Health Promot*. 2000; 14:222-228.
- [8] Serrone, C. Mood management and video-game engagement: The importance of user-experience and gender in assessing the psychological effects of video-game play. (2012). *Master's Theses*. Paper 4211.
- [9] Peng, W. (2009). Design and evaluation of a computer game to promote a healthy diet for young adults. *Health communication*, 24(2), 115-127.
- [10] Giabbanelli PJ, Crutzen R. Supporting self-management of obesity using a novel game architecture. *Health Informatics J*. 2014 Feb 19. [doi: 10.1177/1460458214521051]
- [11] Ledoux T, Griffith M, Thompson D, Nguyen N, Watson K, Baranowski J, Buday R, Abdelsamad D, Baranowski T. An educational video game for nutrition of young people: Theory and design. *Simul Gaming*. 2016 Aug;47(4):490-516. Epub 2016 Feb 17. (full text delayed)
- [12] Lister C, West JH, Cannon B, Sax T, Brodegar D. (2014). Just a Fad? Gamification in Health and Fitness Apps. *JMIR Serious Games*, 2 (2), e9.
- [13] Baranowski T, Baranowski J, Thompson D, Buday R. Behavioral science in video games for children's diet and physical activity change: key research needs. *J Diabetes Sci Technol*. 2011 Mar 1;5(2):229-33.
- [14] Thompson D. Designing serious video games for health behavior change: current status and future directions. *J Diabetes Sci Technol*. 2012 Jul 1;6(4):807-11.
- [15] Owen, V.E., Ramirez, D., Salmon, A. & Halverson, R. (2014). *Capturing Learner Trajectories in Educational Games through ADAGE (Assessment Data Aggregator for Game Environments): A Click-Stream Data Framework for Assessment of Learning in Play*. Presented at the 2014 American Educational Research Assn. Annual Meeting.
- [16] Riet JV, Alblas E, Crutzen R, Lu AS. The Effects of Active Videogames on BMI Among Young People: A Meta-Analysis In book: *Handbook of Research on Holistic Perspectives in Gamification for Clinical Practice*, Publisher: IGI Global, Editors: Daniel Novak, Bengisu Tulu, Havar Brendryen, 2016: pp.277-292. DOI: 10.4018/978-1-4666-9522-1.ch013.
- [17] Spook J, Paulussen T, Kok G, van Empelen P. Evaluation of a Serious Self-Regulation Game Intervention for Overweight-Related Behaviors ("Balance It"): A Pilot Study. *J Med Internet Res* 2016;18(9):e225
- [18] Foots, B. and Roch, J. (2014). The Impact of fooya - a nutrition education mobile health game - on food choices in middle school residential campers at the 2014 ExxonMobil Bernard Harris Summer Science Camp.
- [19] Takahashi, D. (2012). FriendsLearn teaches you through social and mobile games. Retrieved from <http://venturebeat.com/2012/04/18/friendslearn-teaches-you-through-social-and-mobile-games/>
- [20] Sri Prakash, B. (2017). Developing a Digital Vaccine Technology based on Neuropsychology and Cognitive Behavior Therapy to reduce the risk of lifestyle diseases, 2017 Stanford School of Medicine, MedX | ED, Engagement Learning Design: <https://medicinex.stanford.edu/medicine-x-ed-2017-preliminary-program/>

Address for correspondence

Rema Padman, The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA 15213; rpadman@cmu.edu

Acceptance of Using an Ecosystem of Mobile Apps for Use in Diabetes Clinic for Self-Management of Gestational Diabetes Mellitus

Sarita Pais^a, Dave Parry^a, Krassie Petrova^a, Janet Rowan^b

^a School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, New Zealand

^b National Women's Health, Auckland District Health Board, Auckland, New Zealand

Abstract

Mobile applications (apps) for self-management of diseases such as diabetes and for general well-being, including keeping track of food, diet, and exercise, are widely available. However, consumers face a flood of new mobile apps in the app stores and have no guidance from clinicians about choosing the appropriate app. As much as clinicians would like to support a patient-centered approach and promote health and wellness mobile apps, they may be unable to provide advice due to the lack of comprehensive and reliable app reviews. This research reviewed a selection of health and wellness mobile apps suitable for the self-management of gestational diabetes mellitus (GDM). A prototype of an ecosystem that integrated the data generated by the apps was built and its usefulness and ease of use were evaluated. The results show that the ecosystem can provide support for GDM self-management by sharing health and wellness data across the diabetes clinic.

Keywords:

Ecosystem; Diabetes Mellitus; Mobile Applications

Introduction and Background

In recent years the Diabetes Clinic at Auckland Hospital has had an overwhelming number of female patients with gestational diabetes mellitus (GDM) [1]. The traditional method of manually recording blood glucose readings and maintaining a food diary is not efficient when compared to the telehealth technologies available today. For example, women are given glucometers and asked to check their blood glucose readings at least four times a day. The glucometer readings can be downloaded to a computer. Newer glucometers can download the glucometer readings to a smart phone using near field communication (NFC).

The currently available mobile applications (apps) for self-managing blood glucose and diet present another possibility. Although women mostly maintain their food diaries on paper, some women prefer to save them electronically in a spreadsheet and are familiar with mobile apps that monitor exercise and calorie intake. However, the clinicians at the clinic had not trialed the mobile apps that may be suitable for GDM self-management. They were not sure which health and wellness apps may be appropriate and felt they were slow in adopting the new technology. This study describes the development of an extensible "ecosystem" of mobile apps that supports GDM self-management by integrating patient data exported by the apps into the ecosystems' database. Clinicians can use the ecosystem to recommend a suitable app (or apps) to their patients and review the shared data in a clinical

context. More specifically, the research focuses on the app selection process and on the evaluation of the ecosystem's usefulness and ease of use.

Health and Wellness Mobile Apps

There are around 8000 health and wellness mobile apps for managing exercise, healthy living, weight loss, and chronic diseases, available on the two dominant smart phone platforms iPhone and Android [2]. In addition to blood glucose and weight tracking, the common set of functionalities include insulin and carbohydrate tracking, exercise tracking, maintaining a food diary supported by an internal food database, and the ability to share wellness data electronically [3, 4]. For example, My Meal Mate is a smart phone app that allows users to save favorite food combinations and recently logged food entries and to take photographs of food for memory recall [5]. Other useful functionalities include displaying a graph of calories consumed daily and showing the results of the analyses of important macronutrients. In a pilot randomized controlled trial for weight loss it was observed that My Meal Mate had a greater acceptance and satisfaction rating compared to traditional interventions such as using paper based food diaries or online weight loss programs [6].

Mobile App Selection Criteria

It may be difficult for consumers to identify the most suitable app. They can consider the set of features identified above or refer to a review such as Hickman and Elsworth's [7]. These authors identify the advantages and disadvantages of health related mobile apps based on a comparison across a range of criteria including cost, app rating by the public, exercise and food data functionality, the ability to share information on social networks, and compatibility with various mobile phone models. However, such reviews quickly become outdated as new and improved apps are added to the app stores.

The number of downloads and installations of an app gives a measure of its popularity. This indicator may be considered as a useful selection criterion, especially if coupled with checking the app reviews published on the app store web site. However, these reviews are not professional and may lack credibility.

For some consumer categories, interface usability may be a very important aspect of the app's functionality. For example, apps designed for the elderly should have a minimized set of functionalities [8, 9]. The interface should be self-explanatory and convenient to interact with, such as the large buttons used in the eCAALYX (Enhanced Complete Ambient Assisted Living Experiment) app for older people suffering from chronic diseases [8]. Unfortunately, not all app reviews

consider usability features [10]. Similarly, reviews rarely consider the possibility of integrating patient managed data with electronic health records. Ultimately, consumers have to rely on their own judgement in order to identify, trial, and adopt a suitable app.

Sharing Health and Wellness Data with Clinicians

Most commercially available health and wellness mobile apps have a provision to share generated data by email. A clinician can review the electronically shared data during a subsequent consultation with their patient. However, the data are not readily preserved and stored for future consultations, especially with other clinicians. In followup consultations, the physician, the nurse, or the dietician will have no access to the regular updates provided by the app. Hence, it is desirable to store patient managed health and wellness data in a central system accessible to all members of the clinical team. For example, the free to download iPhone app Easy Diet Diary allows for sharing data with a dietician provided that the dietician has purchased a license for Foodworks, a nutrient analysis software [11]. Easy Diet Diary uses food databases specific to Australia and New Zealand.

Exporting patient managed data to a proprietary database does not allow patients the flexibility to select the app of their choice. Furthermore, it may be convenient for patients to manage multiple health conditions through one health and wellness mobile app. Unfortunately, clinical trials and reviews of commercially available apps provide limited evidence about integrating health and wellness data with clinical systems.

This research set out to create a prototype of an ecosystem of mobile apps and devices that can be used to support GDM self-management. The system accepts, stores, and integrates shared health and wellness data from the ecosystem's components and makes them available to clinicians to search and review (Figure 1). The study addresses the following two research questions related to the design and implementation of the ecosystem:

RQ1. What are clinician criteria for including health and wellness apps in the ecosystem?

RQ2. What are the perceived usefulness and the perceived ease of use of the ecosystem?

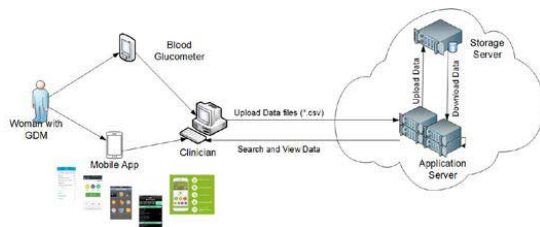


Figure 1 - An Ecosystem Integrating Health and Wellness Data

Method

A user-centered design approach towards the development of the ecosystem prototype was chosen as it needed to meet the requirements of its potential users, i.e., the clinicians treating women with GDM at the Diabetes Clinic and their patients. Information about the functional requirements and the data elements needed to build the ecosystem was obtained from physicians, obstetricians, dieticians, and midwives who

reviewed and evaluated mobile apps with food diary and exercise logging functionalities. Data were gathered through open-ended interviews [12]. The data were analyzed qualitatively in order to identify clinician criteria.

Participants

Two dieticians, a midwife, an obstetrician, and a physician participated in an initial mobile app evaluation round that aimed to extract design requirements. The ecosystem prototype was evaluated for its perceived usefulness and ease of use by the initial five participants, by five other clinicians from the same clinic, and by five patients (women with GDM). No real patient data were stored in the prototype.

Prototype Development

The ecosystem prototype was developed in stages with feedback sought from clinicians. The first round of interviews (with five clinicians) contributed to formulating the user requirements of the prototype. Clinicians were periodically emailed a stakeholder consultation document that provided details about the prototype as it was developed. This approach was chosen because clinician availability was limited, making focus group meetings difficult to organize and convene.

The stakeholder consultation document contained descriptions of the low fidelity prototype interfaces, the navigation features, and the prototype data elements necessary for GDM self-management. It also included a review of health and wellness apps that could potentially be added to the ecosystem. The clinicians trialed the apps specified in the document and provided feedback in a continuous dialogue conducted through email, text messages, and phone calls. As the research progressed, other mobile apps were suggested by the participants and more were also discovered by the researchers.

Results

The stakeholder document initially included samples of the data exported from five health and wellness mobile apps selected by the researchers: My Meal Mate, Glucose Buddy, On Track, Doctor Diet, and Microsoft HealthVault. Sending the initial stakeholder document to clinicians started the process of iteratively and interactively capturing and refining the clinician criteria. Three selection criteria were used. First, all apps needed to support food diary and exercise functionality. Second, all apps needed to allow data sharing with third parties. Third, based on findings that cost was a significant factor influencing consumer's choice of a mobile app, [13] only free to download apps were considered.

Table 1 shows a summary of the properties of the first five health and wellness mobile apps reviewed by the research participants. All apps supported food diary functionality while four apps supported exercise tracking. However, the absence of a food database, as in the case of the apps OnTrack and Glucose Buddy, would make data entry rather difficult (and potentially inaccurate) as patients would have to fill in all food details.

Four of the apps provided options for exporting data that could be shared with clinicians. The app My Meal Mate did not have a data sharing option. However, the food and exercise data were stored in the user's mobile phone in a SQLite database. A reverse engineering process that used the app's internal food database allowed for extraction of data about food and exercise that could be integrated in the ecosystem and shared

with clinicians. The data also provided nutrient information about the food consumed. However, the app's food database and nutrient profiles were specific to branded food items found in UK supermarkets and not relevant to New Zealand app users.

Table 1 - Health and Wellness Mobile App Review

App	Food diary	Exercise	Export data to clinician	Data Format
My Meal Mate	Yes	Yes	No	SQLite
Glucose Buddy	Yes	Yes	Yes	CSV
OnTrack	Yes	Yes	Yes	XML
Doctor Diet	Yes	No	Yes	CSV, pdf report
Microsoft Health Vault	Yes	Yes	Yes	CSV

Microsoft HealthVault was available as an app for iPhone, but there was no equivalent suitable for Android phones. By default, the apps' capability to share data about food details with a clinician or a friend was limited. The settings can be changed to allow greater food data sharing. However, the process is not user friendly; it may be difficult for an ordinary user to perform the manipulation.

A drawback of the Doctor Diet app was that it focused on food and diet only. Users would have to install another app in order to record exercise.

The app Easy Diet Diary was added to the ecosystem in a subsequent stakeholder document. It features an easy to enter food dairy interface and supports extracting data for later integration within the ecosystem. However, as already mentioned, a clinician would need to purchase a license for the Foodworks app, in order to enable data sharing.

Clinician Criteria

Clinicians reviewed the apps' properties and functionalities. Their views and opinions were used to derive a set of criteria that mobile health and wellness apps needed to meet in order to be considered for inclusion in the ecosystem.

First, in order to be relevant to GDM self-management, the app needed to provide data about both physical activity and food intake, and also data about glucose level and treatment.

Clinician 4

Reflects physical activity and food diary together with treatment and dosage and sugar level.

Second, it was established that a mere calculation of the calories consumed by the patient was not sufficient as the intention of the app user, in this case, was not to lose weight but to manage their diet in order to control GDM. The type of food, its description, and portion size were important indications that needed to be captured, especially for dieticians and midwives. Such data would help clinicians support women with GDM to interpret glucose readings and manage insulin dosage.

Clinician 1

Yes, it is including pretty much everything that you want to know about: blood glucose, food diary, exercise. Calories will not tell you what you have eaten. Need to choose apps which

are usable. You need app(s) which get the calories, carbohydrate breakdown, protein and fat.

However, an app's capability to record carbohydrate content or carbohydrate counting was not critical as most women had GDM only during pregnancy. Furthermore, such patients may not have the carbohydrate counting skills commonly acquired by women with ongoing Type 1 diabetes. Therefore, the clinic had to develop a different approach when consulting women with GDM. In general, apps focusing on healthy lifestyle support would be suitable and more appropriate for women with GDM.

Clinician 3

Wellness data is representative. Easy to compare food diary, exercise, insulin dosage in comparison with blood glucose.

Only two apps, Glucose Buddy and On Track, satisfied all requirements. These apps allowed food description with portion size, which was required by dieticians and midwives. Thus, the ecosystem that was built following the initial requirement analysis included obtaining and integrating patient health and wellness data generated by Glucose Buddy and On Track. The blood glucose readings from the glucometer were also integrated into the prototype.

Ecosystem Usefulness and Ease of Use

The data extracted from the apps and available in the prototype's database were reviewed by participants in order to obtain further insights into the suitability of the apps. It was found that the perceived usefulness of the ecosystem was directly related to the perceived usefulness of the data provided by the mobile apps.

Clinicians reviewing the ecosystem found it useful as it succeeded in bringing together, in a single report, data from blood glucose readings, the food diary, and the exercise tracking.

Most of the women with GDM who participated in the research had experience using mobile apps, for example managing their diet, weight, and exercise (step counter). However, they were not knowledgeable about mobile apps that may be suitable for GDM self-management. After reviewing the mobile apps included in the ecosystem, women with GDM became interested in using them. They were enthusiastic about sharing their wellness data with clinicians as it would help them self-manage the condition, benefiting their own health and that of their babies.

Patient 1

(I would) absolutely use it and comfortable to share the data with clinician if suitable apps are available.

Both clinicians and women with GDM found the prototype easy to use as completing a task did not require going through many screen interfaces. For example, viewing a combined report about a food diary, blood glucose readings, exercise, and insulin dosage required, in most cases, one screen.

Most women preferred to log their food entries electronically rather than maintain a paper based record. However, training may be required for some, such as first time smart phone owners. Similarly, clinicians were confident they would learn to use the system as it resembled other hospital systems and would be used often (each time a patient visited the clinic). The study findings about the dimensions of perceived usefulness and ease of use of the ecosystem are presented in Table 2.

It took almost a year to gather the initial requirements, build the prototype, and review the ecosystem. At the end of the

process, most women visiting the clinic owned smart phones and were aware of or were already using health and wellness mobile apps. They often asked clinicians to recommend a suitable app. This highlighted yet again the usefulness of the ecosystem and the need to encourage and stimulate the adoption of the new technology by clinicians.

Table 2 - *Ecosystem usefulness and ease of use*

Variable	Study Findings
Perceived Usefulness	<ul style="list-style-type: none"> • Sharing patient data with the clinician team • Combining data from various sources • Empowering patients in self-management • Helping in food recall • Remotely managing patients • Reducing cost
Perceived Ease of Use	<ul style="list-style-type: none"> • Condensing data in one screen • Relevant information in one screen • Easy navigation • Existing experience with /awareness of mobile apps • Minimal training required

Extending the Ecosystem

Data interoperability was achieved by mapping the schema of the data coming from a mobile app to the target database in the ecosystem. The ecosystem design allowed the easy inclusion of new apps. Patient participants identified new apps that could be added to the ecosystem in order to increase its scope, e.g., MyFitnessPal and mySugr. Both apps were free to download and available for the most popular smart phones, iPhone and Android. Data generated by the apps could be shared by email.

mySugr had the option to export data in CSV format and could be easily incorporated in the ecosystem, thus extending its range of available apps. MyFitnessPal had the option to generate a report in pdf format. The “premium” (paid for) version of MyFitnessPal allowed sharing collected data with other partnering vendors such as FitBit and JawBone using a private application program interface (API). Although not useful for users with a non-premium account, MyFitnessPal is very popular. If its owners add the data sharing feature to the free version, adding the app to the ecosystem may be considered.

Another potential extension was the already assessed Easy Diet Diary app. It exports data to a proprietary nutrition database (Foodworks). However, its usefulness is rather limited as it is available for iPhone only.

Reviews about health and wellness mobile apps by members of the health care community provided a useful means for identifying more apps that could potentially be added to the ecosystem. For example, a group of dieticians from various hospitals and clinics review apps suitable for self-management of chronic diseases and general well-being. The reviews are published in a web portal accessible to the general public [14]. The reviews consider apps from different user perspectives (i.e., clinician or patient) and also from a technical perspective. The web site contains a list of mobile apps supporting healthy living and managing food, diet, and exercise. The list includes the already mentioned MyFitnessPal and mySugr, and also a paid app called Glooko.

In order to be able to integrate shared data from Glooko in to a central database, the clinic would also need to purchase a license. However, clinics with limited funding may not be able to invest in adding paid apps to the ecosystem unless the benefits in terms of self-management interventions and health outcomes are shown to be of much higher value compared to free apps with similar functional capabilities.

Overall, the prototype showed that an extensible ecosystem of mobile apps suitable for the self-management of GDM can be successfully set up. To support privacy protection, the ecosystem facilitated patient access to their health and wellness data while clinicians had to be authorized by patients in order to access their data. Such practices are recommended by researchers [15, 16].

Discussion

The ecosystem described above helped demonstrate the potential of using mobile apps in the self-management of GDM. New and improved apps may evolve. With appropriate clinician approval and evaluation, these may be ready for integration into the ecosystem. Evaluation criteria can be developed using the required app features (Table 1) and the usefulness and ease of use dimensions identified in the research (Table 2).

The selected Android and iPhone apps had the capability of exporting data in various formats, either shared through an email or saved in a cloud database. In the first version of the prototype, clinicians suggested that they should upload patient data to the ecosystem’s database themselves. They did not trust the accuracy of the data created by patients and wanted to review them first. Clinicians preferred receiving data through email. In most cases, data in formats such as CSV and XML would open easily without installing additional software. However, the next step (integrating the data in the prototype’s database manually) appeared to be an additional workload for clinicians, especially for midwives who met women with GDM on a regular basis.

Manually managing each patient’s file is not an efficient method of exchanging health data. There is evidence of better health outcomes when patients are involved in the self-management of their own diseases [17, 18]. Patients should be trusted to upload their own wellness data into the prototype accurately, even if patient generated health data are somewhat dissimilar to data captured by clinicians and stored in conventional health information systems. Sharing patient generated data across the clinic is beneficial as it helps clinicians keep track of their patients’ progress remotely and in a cost efficient manner.

Data entry and exporting data from mobile apps to another system is not yet automated and seamless. However, existing APIs have the potential to share patient generated health and wellness data from participating mobile apps and sensor devices, thus enabling the creation of an app ecosystem. Data interoperability is achieved as APIs “know” the source and target data schema. However, it can be expected that future mobile apps will have an API capability of transparently sharing data with other systems. For example, several major commercially available wearable tracking devices for step counting, sleep, and heart beat monitoring already support sharing data with other apps.

Although data interoperability was achieved by mapping data from mobile apps to the target prototype’s database, semantic interoperability is still an issue. Clinical standards such as

SNOMED CT, HL7 and the recently introduced draft version of FHIR do not have the provision to include patient generated health and wellness data.

Conclusion

An ecosystem of mobile apps for GDM self-management was set up for patients seen in Diabetes Clinic at Auckland Hospital. Clinicians reviewed the requirements of the ecosystem through a continuous dialogue and discussion by email, phone conversations, and interviews. The review process helped formulate app selection criteria and identify the usefulness and ease of use dimensions of the ecosystem. The women with GDM who participated in the research were confident that the use of these new technologies would help them.

Promoting health information exchange and system interoperability is a prime objective of the “meaningful use” data sharing framework [19]. Despite potential changes in this, it is likely that in the future, services allowing secure, accurate, and timely health data exchange between consumer-focused systems will be available through APIs. This study contributes to the body of knowledge in the area of electronic sharing of health and wellness data from consumer wellness systems in order to improve the quality of health care and involve citizens as active participants in the health care process.

Acknowledgements

The authors thank all participants in the study who provided the useful feedback that informed the research.

References

- [1] National Women's Annual Clinical Report. <http://nationalwomenshealth.adhb.govt.nz/Portals/0/Annual%20Reports/ACR%20master%20MOH%20July%202015.pdf>
- [2] Aitken M. Patient Apps for Improved Healthcare: From Novelty to Mainstream. IMS Institute for Healthcare Informatics. http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Corporate/IMS%20Health%20Institute/Reports/Patient_Apps/IIHI_Patient_Apps_Report.pdf 2013.
- [3] Tran J, Tran R and White JR. Smartphone-based glucose monitors and applications in the management of diabetes: an overview of 10 salient “apps” and a novel smartphone-connected blood glucose monitor. *Clinical Diabetes*. 2012; 30(4): 173-178.
- [4] Eng DS and Lee JM. The promise and peril of mobile health applications for diabetes and endocrinology. *Pediatric diabetes*. 2013; 14(4): 231-238.
- [5] Carter MC, Burley VJ, Nykjaer C and Cade JE. ‘My Meal Mate’ (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. *British Journal of Nutrition*. 2013; 109(03): 539-546.
- [6] Carter MC, Burley VJ, Nykjaer C and Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. *Journal of medical Internet research*. 2013; 15(4).
- [7] Hickman I and Elsworth B. Review of apps for monitoring food intake and exercise. <http://www.dietitianconnection.com/component/content/article/52-products/resources-materials/924-review-of-apps-for-monitoring-food-intake-and-exercise> 2012.
- [8] Boulos MN, Wheeler S, Tavares C and Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX”. *Biomedical engineering online*. 2011; 10(24): 1-14.
- [9] Prescher S, Bourke AK, Koehler F, Martins A, Sereno F H, Boldt ST and Nelson J. Ubiquitous ambient assisted living solution to promote safer independent living in older adults suffering from co-morbidity, *In*

Conference proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011: 5118-5121.

- [10] El-Gayar O, Timsina P, Nawar N and Eid W. Mobile applications for diabetes self-management: status and potential. *Journal of diabetes science and technology*. 2013; 7(1): 247-262.
- [11] Easy Diet Diary. <https://easydietdiary.zendesk.com/hc/en-us/articles/206615603-How-do-I-send-my-diary-to-my-nutrition-professional-to-view-in-FoodWorks->
- [12] Martin JL, Clark DJ, Morgan SP, Crowe JA and Murphy E. A user-centred approach to requirements elicitation in medical device development: A case study from an industry perspective, *Applied Ergonomics*, 2012; 43(1): 184-190.
- [13] Krebs P and Duncan DT. Health app use among US mobile phone owners: a national survey. *JMIR mHealth and uHealth*. 2015; 3(4).
- [14] Health Navigator. <https://www.healthnavigator.org.nz/app-library/d/diabetes-apps/>
- [15] JASON, Mitre Corporation. A robust health data infrastructure. US Agency for Healthcare Research and Quality publication no. 14-0041-EF. http://healthit.gov/sites/default/files/ptp13-700hhs_white.pdf. 2014.
- [16] Hull S. Patient-Generated Health Data Foundation for Personalized Collaborative Care. *Computers Informatics Nursing*. 2015; 33(5): 177-180.
- [17] Holtz B and Lauckner C. Diabetes management via mobile phones: a systematic review”. *Telemedicine and e-Health*, 2012; 18(3): 175-184.
- [18] Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P and Haines A. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review”. *PLoS Med*. 2013; 10(1): e1001362.
- [19] Medicare and Medicaid Programs. Electronic Health Record Incentive Program-Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. <https://www.federalregister.gov/documents/2015/10/16/2015-25595/medicare-and-medicaid-programs-electronic-health-record-incentive-program-stage-3-and-modifications>

Address for correspondence

Sarita Pais, sarita.pais@aut.ac.nz

+64 21 0379273 Alternate email pais.sarita@gmail.com

Establishing Therapeutic Alliance in Mental Health Care via Cooperative Documentation

Anja Perlich, Julia von Thienen, Christoph Meinel

Hasso Plattner Institute, University of Potsdam, Germany

Abstract

In talk-based mental health interventions, treatment outcomes can be decisively improved by enhancing the relationship between patient and therapist. We developed the interactive documentation system Tele-Board MED (TBM) with the goal of supporting patients and doctors in their cooperative task of patient care. The system offers a whiteboard-inspired graphical user interface which allows them to take notes jointly during the treatment session. Two proxy studies were conducted whereby TBM was introduced in a role play that showcased the dialogue in a therapy session. The patient role was played by a volunteer. The audience of human-centered design as well as eHealth experts rated the therapist-patient relationship in a session with and without TBM. The data collected via questionnaires shows that TBM consistently receives a positive rating from study participants (N=36) in the areas of collaboration, communication, patient-doctor relationship, as well as patient empowerment.

Keywords:

Physician-Patient Relations; Patient Participation; Psychotherapy

Introduction

Mental health disorders account for over 40% of all chronic illnesses and are the biggest cause of years lived with disability in the developed countries [1]. One major reason why treatments fail is a lack of engagement by the patients. Recent studies show that technology has the potential of supporting a productive patient-doctor relationship and of promoting patient engagement [2]. However, the use of information and communication technology in talk-based mental health interventions is fairly limited, to a large degree because research in this domain is restricted by sensitivity and social stigma. Symptoms of mental disorders may include anxiety, depressed mood, and obsessive-compulsive or delusional behaviour. Talk-based interventions for diagnostics and treatments are common approaches.

In the domain of mental health care, the therapist-patient relationship is a primary curative component and can be considered a necessary prerequisite for the effectiveness of all therapeutic interventions [3]. The term ‘therapeutic alliance’ is used to describe the ideal patient-provider relationship. This alliance is defined as “a dynamic interactional process in which the patient and provider collaborate to carry out negotiated mutual goals in a shared partnership” [4].

In this paper we elaborate on the support of patient-doctor relationships in psychotherapeutic treatments through the use of cooperative technology. We introduce the documentation system Tele-Board MED (TBM), which aims at supporting

doctor-patient interactions in face-to-face clinical encounters. Traditionally, the therapist takes handwritten notes (figure 1a) in order to serve personal purposes, e.g. to fulfil legal duties and to overview the treatment. Inviting patients to access their record is very unlikely, as its legibility and understandability can hardly be guaranteed [5]. In contrast, TBM allows the patient and the care provider to jointly take notes and freely structure them on a whiteboard-inspired graphical user interface (figure 1b). The documentation panels can be filled with digital sticky notes, uploaded images, and visual elements such as scribbles. This way, progress notes, therapy material, and case stories are collected in a digital, visually enhanced patient file. Blank documentation panels can be used, as well as templates for specific tasks or treatment approaches (see figure 2).



Figure 1a – Traditional treatment session: Therapy notes are visible to the therapist only.



Figure 1b – Treatment session with Tele-Board MED: Therapist and patient work together on therapy notes by using a digital whiteboard, a wireless keyboard and a tablet computer.



Figure 2a – Example of a documentation panel used in anamnesis sessions containing headlines and patient information visible as sticky notes.

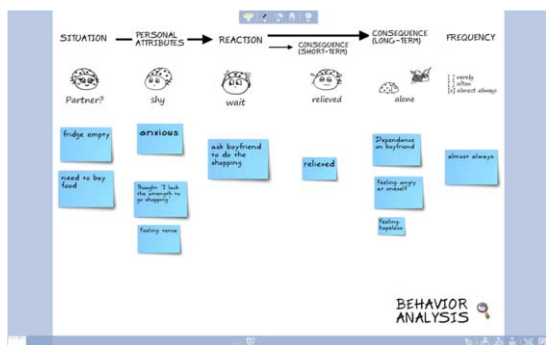


Figure 2b – Example of a documentation panel used for behavior analysis based on the SORKC model [6].

Related Work

Tele-Board MED is based on Tele-Board – a digital whiteboard system to support creative teamwork over distances [7]. It is a web application based on HTML5 technology [8], which runs in a web browser and thus on diverse hardware devices, such as a desktop computer, laptop, tablet computer, or interactive whiteboard.

Technology used in mental health care to date strongly focuses on improving the access to services and mostly represents a replication of existing, analogue methods. There are means of electronic communication, such as email or chat to support doctor and patient collaboration. Furthermore, there are technologies aimed at increasing patient engagement. Classic therapy techniques, such as emotion regulation and exposure, can be enhanced by virtual or augmented reality [9, 10]. In addition to technology interventions introduced by the therapist, there are computerized therapy applications that autonomously deliver care, e.g. internet-based interventions for post-traumatic stress disorders [11]. However, in order to not leave patients alone with their problems, Knowles et al. [12] call for more research about computer-mediated therapy looking at the integration of health professional input with computer-delivered content.

Tele-Board MED, as well as the software applications described in the following, fall in the category of computer-mediated therapy. Coyle and Doherty [2] designed the role-playing computer game ‘Personal Investigator’ to be played in therapy sessions by an adolescent patient together with a therapist, who takes the role of a partner in the exploration of a game world. A digital notebook supports the player’s personal reflection. In contrast, TBM strives for a collaborative documentation by both patient and therapist. The mobile phone and web application ‘My Mobile Story’ incorporates the therapeutic agent of telling and re-telling the patient’s story to trigger self-reflection. This is carried out by capturing case information in visual and multimedia-based ways [13]. In contrast to TBM, it seems that collaboration between therapist and patient via the application takes place asynchronously and remotely.

Research Questions

Our research objective with TBM is to increase patient engagement in treatments and to improve the doctor-patient relationship. In this way, the effectiveness of the overall treatment increases. The described study addresses the following questions:

- Can Tele-Board MED help to strengthen the therapeutic alliance between patient and therapist?
- Can Tele-Board MED help to increase the shared knowledge of patient and therapist?

In the following, we describe the study setup, the collection of quantitative data through questionnaires, and the data analysis in comparing a traditional therapy session scenario with a TBM-supported scenario. We discuss our study design and results in depth – and finally conclude with a summary.

Methods

In order to test the effects of Tele-Board MED on therapeutic alliance and patient empowerment, two proxy studies [14] were conducted.

Study Setup

In both studies, TBM was presented in the form of an impromptu role play that showcased a therapeutic dialogue. The first study was conducted in Germany with 8 participants whose professional background was eHealth. We used a digital whiteboard as hardware device. The second study was conducted in the United States with 28 participants from the field of human-centered design. Here, a laptop and projector were used. The role of the psychotherapist was acted by a mental health care professional. In both studies, a person from the audience spontaneously volunteered to act the patient role and shared a personal problem. A short, true-to-life therapy session was carried out on stage.

In both settings, the role play contained two parts, each of which was followed by data collection (see figure 3). First, the therapist used traditional documentation means: handwriting on paper sheets. After about five minutes, a break was taken. Both the audience and the proxy patient were asked to fill out short questionnaires. Afterwards, during the next five minutes of therapeutic conversation, TBM was used as a means of documentation. Then again, audience and “patient” filled out questionnaires.

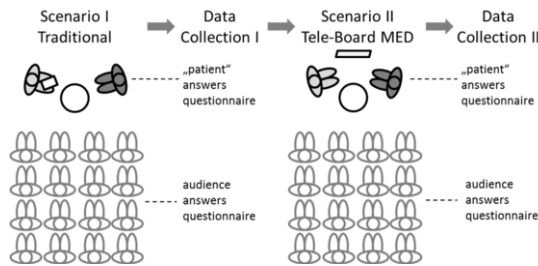


Figure 3 – Procedure in both proxy studies: the role play contained two parts, each of which was followed by data collection.

Questionnaires for Data Collection

The questionnaire items represent several aspects of relevance for patient-doctor collaboration (see table 1). The constructs of collaboration, integration, empowerment and communication represent the four subscales of the Kim Alliance Scale [4]. In each subscale one or two of the corresponding items were adopted. The items in table 1 are phrased from the patient perspective. For the audience questionnaire, the syntax of the sentences was adapted (e.g. item 2 reads “I feel my therapist listens to me.” for the patient questionnaire and “The therapist listens to the patient.” for the audience questionnaire). The items were rated on a five-point scale, ranging from 2 to -2 (‘clearly so’{2}, ‘seems so’{1}, ‘I don’t know’{0}, ‘doesn’t seem so’{-1} and ‘clearly not’{-2}). Feedback from the first study led to minor modifications in the questionnaire (one item was omitted, two new items were added). Thus, most items were rated by all study participants (N=36), however, item 5 was rated by 8 people and items 8 and 9 were rated by 28 people.

Table 1 – Items used to assess patient-therapist cooperation with and without Tele-Board MED. The constructs marked with an (*) are the dimensions of the Kim Alliance Scale [4].

Construct	#	Item
Collaboration*	1	Me and my therapist have the same therapeutic goals.
Integration*	2	I feel my therapist listens to me.
Empowerment*	3	I am allowed in the decision-making process.
Communication*	4	It is easy to understand my therapist’s instructions.
	5	My therapist and I work well together.
Relationship	6	Me and my therapist collaborate at eye-level.
Documentation	7	It is possible for me to recognize documentation errors.
Shared Knowledge	8	Me and my therapist develop joint knowledge that we can build on in the next session.
	9	Me and my therapist have a common understanding of the treatment procedure.

Statistical Measures

For the statistical analysis of the quantitative questionnaire data, the software package SPSS Version 22 was used. We calculated the arithmetic means of both scenarios (traditional

vs. TBM), as well as measures to assess the statistical significance (two-sided p-values) and the effect size (Cohen’s d). In order to determine the degree of agreement among the audience raters, we calculated intra-class correlation coefficients for single measures and Cronbach’s alpha for average measures. Coefficients of Pearson, Spearman’s rho and Kendall’s tau-b were used to measure pairwise correlation among the two proxy patient raters.

Results

The data of both proxy studies was analyzed together. The following sections describe the comparison of both scenarios and provide assessments of statistical significance and reliability.

Evaluation: Traditional vs. Tele-Board MED

Table 2 shows the arithmetic means (in a range from 2 to -2) of the ratings for the traditional and the TBM scenario pertaining to the audience and the volunteering “patients”. For the audience ratings, dependent t-tests for paired samples with two-tailed tests of significance (p-values) show the level of statistical significance. These tests were chosen because each person provided ratings for the traditional and the TBM setting, and because there was no a priori knowledge about which scenario would receive better ratings. Furthermore, Cohen’s d estimates the effect size of the intervention. Here, the traditional setting is interpreted as the control condition and the TBM setting is considered the experimental condition, so that positive d-values indicate a positive effect of TBM. N specifies the number of valid answers per item. Due to some missing replies, and three items not being part of both studies, N varies across the comparisons.

For the answers of the audience members who volunteered to be patients, no statistical tests are computed since there are only two respondents.

Table 2 – Average ratings of questionnaire items (#) for a traditional therapy setting (Trad.) versus a session with Tele-Board MED (TBM) listed separately for audience members and volunteering patients. N specifies the number of comparisons. A two-sided p-value (p) indicates the level of statistical significance. Cohen’s d (d) indicates the effect size.

#	AUDIENCE					VOLUNTEER		
	Trad.	TBM	N	p	d	Trad.	TBM	N
1	0.06	0.97	32	<.001	1.25	0.5	0.5	2
2	1.24	1.15	33	.572	-0.12	1.5	2	2
3	-0.58	1.42	31	<.001	2.62	-5	1.5	2
4	0.42	1.23	31	<.001	1.03	2	2	2
5	0.86	1.57	7	.008	1.15	1	2	1
6	0.03	0.81	32	.010	0.77	-0.5	2	2
7	-1.64	1.64	33	<.001	5.44	-1.5	1.5	2
8	-0.44	1.28	25	<.001	2.13	0	1	1
9	-0.52	0.60	25	<.001	1.46	-1	0	1

Inter-Rater Reliability

For the ratings by the audience members, intra-class correlation coefficients for single measures and Cronbach’s alpha for average measures were calculated.

In the first study with audience ratings of eHealth experts in Germany (N=7), the intra-class correlation coefficient, calculated with a two-way random model and consistency

analysis, amounts to 0.69 for single measures (i.e. regarding single items). On a level of $p < 0.001$, this is statistically significant. In terms of average measures (i.e. regarding the total amount of items), Cronbach's alpha is 0.94, where 1 would be the maximum possible. When this analysis is carried out in the second study with audience ratings of human-centered design experts in the United States ($N=27$), the intra-class correlation yields a coefficient of 0.870 for single measures, which is also significant on a level of $p < 0.001$. Here, Cronbach's alpha for average measures is 0.995. Considering the audience raters of study one and two together ($N=34$) and only those 12 items (2x6) that were handed out in both studies, the intra-class correlation coefficient amounts to 0.873 for single measures, which is again significant on a level of $p < 0.001$. In terms of average measures, a Cronbach's alpha of 0.996 is obtained.

To assess the inter-rater agreement of the proxy patients, common correlation coefficients were calculated. On metrics ranging from 1 (a perfect positive relationship) to 0 (neutral relationship) and down to -1 (a perfect negative relationship), all measures yield values above 0.8, which is statistically highly significant (Spearman's rho: 0.89 with $p < 0.001$, Pearson: 0.86 with $p < 0.001$, Kendall's tau-b: 0.81 with $p < 0.002$).

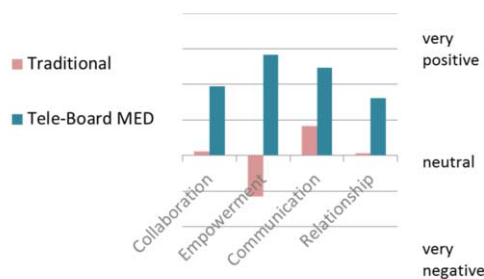


Figure 4 – Expert ratings ($N=36$) on patient-doctor cooperation indicate great benefits of Tele-Board MED compared to traditional documentation approaches.

Discussion

The use of Tele-Board MED seems to have highly significant positive effects on therapist-patient interactions. Figure 4 displays a chart with the expert ratings regarding the comparison of the traditional versus TBM scenario based on the numbers of table 2 relating to the four items and respective constructs #1 (collaboration), #3 (empowerment), #4 (communication), and #6 (relationship).

The two studies on TBM effects had a differing session content (i.e. each proxy patient introduced an individual problem) and different display devices were in use. The studies were conducted on two different continents and the professional background of the expert audiences varied. These differences are likely to add some variance ("noise") to the data. Still, the ratings show high intra-class correlations both within and across studies. Therefore, the findings seem all the more reliable. A high concordance indicates that the questionnaire answers seem noncontroversial or obvious. Subjective or culture-specific viewpoints bear little influence on the ratings. Consequently, the participants (despite of belonging to different rater populations) show a striking agreement in their responses along all questionnaire items.

There is a strongly significant effect in favor of TBM in eight out of nine items. Statistically, a p -value below 0.05 is generally considered significant and hence it would be sufficient to support the hypothesis that TBM makes a difference. Here, even p -values below 0.001 are observed, which is remarkable given the relatively small sample sizes. This is indicative of a significant effect (see table 2). It is likewise reflected by Cohen's d , where values of 0.2 to 0.5 are considered a small effect, values of 0.5 to 0.8 indicate a medium effect and values above 0.8 indicate a large effect [15]. The questionnaire data does not only indicate a large positive effect on almost all scales. There are even shifts from a negative average rating without TBM to a positive rating with TBM. The only variable where the audience does not see a clear positive effect of TBM is item #2 ("The therapist listens to the patient."). However, the members of the audience who volunteered to take on the patient role do indicate that they experience a positive effect of TBM on this item as well (see table 2, part on the right).

Given that therapeutic alliance is seen as one of the most important factors in predicting patient outcome [3], the concordant results of two proxy studies, carried out on two different continents with two different expert audiences, can be considered very promising. Hence, we conclude that TBM has the potential of delivering more than documentation support, namely strengthening the therapeutic alliance between patient and therapist and increasing their shared knowledge. These findings are in line with qualitative insights from real patient consultation with TBM that showed an increased acceptance of diagnoses and patient-therapist bonding [16].

In future work, experimental assessments in on-site clinical studies should add further data to evaluate TBM.

Conclusion

We introduced the cooperative documentation system Tele-Board MED as a supportive tool for face-to-face encounters in mental health care. In this domain, characterized by sensitivity and stigma, research studies conducted with substitute persons instead of real patients are an increasingly valued methodological approach. In the proxy studies discussed here, we found evidence that therapy documentation can be turned from a necessity taken care of by the doctor into an intervention that affects treatments positively. TBM creates more benefits than providing patients access to their treatment notes. With great concordance, participants observed positive effects of TBM on the creation of shared knowledge and therapeutic alliance between patient and care provider. The latter aspect indicates that TBM strengthens therapeutic communication, integration, collaboration, and patient empowerment.

Acknowledgements

We would like to express our appreciation of the HPI-Stanford Design Thinking Research Program for funding this project, and we thank our clinical partners, our expert audience and volunteers.

References

- [1] World Health Organization (WHO), *Mental health: facing the challenges, building solutions*. World Health Organization, 2005.

- [2] D. Coyle and G. Doherty, Clinical Evaluations and Collaborative Design : Developing new technologies for mental healthcare interventions, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, 2051-2060.
- [3] M.J. Lambert and D.E. Barley, Research summary on the therapeutic relationship and psychotherapy outcome, *Psychotherapy: Theory, Research, Practice, Training* **38**(4) (2001), 357-361.
- [4] S.C. Kim, S. Kim, and D. Boren, The Quality of Therapeutic Alliance between Patient and Provider Predicts General Satisfaction, *Military Medicine* **173**(1) (2008), 85-90.
- [5] A. Perlich and C. Meinel, Automatic Treatment Session Summaries in Psychotherapy – A Step towards Therapist-Patient Cooperation, *Procedia Computer Science* **63** (2015), 276-283.
- [6] F.H. Kanfer and G. Saslow, Behavioral analysis: An alternative to diagnostic classification, *Archives of general psychiatry* **12** (1965), 529-38.
- [7] R. Gumienny, L. Gericke, M. Quasthoff, C. Willems, and C. Meinel, Tele-Board: Enabling efficient collaboration in digital design spaces, *Proceedings of the International Conference on Computer Supported Cooperative Work in Design*, 2011, 47-54.
- [8] M. Wenzel, L. Gericke, R. Gumienny, and C. Meinel, Towards Cross-Platform Collaboration - Transferring Real-Time Groupware To The Browser, *Proceedings of the International Conference on Computer Supported Cooperative Work in Design*, 2013, 49-54.
- [9] O. Baus and S. Bouchard, Moving from virtual reality exposure-based therapy to augmented reality exposure-based therapy: a review, *Frontiers in Human Neuroscience* **8** (2014), 1-15.
- [10] R. Gonçalves, A. L. Pedrozo, E. S. F. Coutinho, I. Figueira, and P. Ventura, Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review, *PLoS One* **7** (2012), 1-7.
- [11] A. Kuester, H. Niemeyer and C. Knaevelsrud, Internet-based interventions for posttraumatic stress: A meta-analysis of randomized controlled trials, *Clinical Psychology Review* **43** (2016), 1-16.
- [12] S. E. Knowles, G. Toms, C. Sanders, P. Bee, K. Lovell, S. Rennick-Egglestone, D. Coyle, C.M. Kennedy, E. Littlewood, D. Kessler, S. Gilbody, P. Bower, Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety, *PLoS One* **9** (2014), 1-12.
- [13] M. Matthews, G. Doherty, and J. Sharry, My Mobile Story Available: <https://www.scss.tcd.ie/misc/TMH/project/my-mobile-story> [Accessed: 25-October-2016].
- [14] J.L. Boyd-Graber, S.S. Nikolova, K.A. Moffatt, K.C. Kin, J.Y. Lee, L.W. Mackey, M.M. Tremaine, and M.M. Klawe, Participatory design with proxies, *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, 151-160
- [15] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*, Lawrence Erlbaum Associates Inc, 1988.
- [16] A. Perlich and C. Meinel, Patient-provider teamwork via cooperative note taking on Tele-Board MED, *Exploring Complexity in Health: An Interdisciplinary Systems Approach*, 2016, 117-121.

Address for correspondence

Anja Perlich

Hasso Plattner Institute (HPI), University of Potsdam

Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany

Phone: +49(0)331 5509 530

Email: anja.perlich@hpi.de

Integrating Social Networks and Remote Patient Monitoring Systems to Disseminate Notifications

Hugo A. Ribeiro, Eliseu Germano, Sergio T. Carvalho, Eduardo S. Albuquerque

Instituto de Informática, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

Abstract

Healthcare workforce shortage can be compensated by using information and communication technologies. Remote patient monitoring systems allow us to identify and communicate complications and anomalies. Integrating social networking services into remote patient monitoring systems enables users to manage their relationships. User defined relationships may be used to disseminate healthcare related notifications. Hence this integration leads to quicker interventions and may reduce hospital readmission rate. As a proof of concept, a module was integrated to a remote patient monitoring platform. A mobile application to manage relationships and receive notifications was also developed.

Keywords:

Social Networking; Patient Monitoring; Physiology

Introduction

Healthcare systems might not properly accommodate all patients in the near future. There is a shortage of healthcare professionals according to World Health Organization [1]: healthcare workforce will be short by 12.9 million professionals in 2035. Also, the world population is demanding more care given the increased life expectancy and the occurrence of non-communicable diseases [2].

Remote patient monitoring systems may reduce the stress on healthcare systems. The stress relief is achieved via the use of technologies that enable communication between patients and health professionals at any time. Using information and communication technologies might also reduce personalized care delivery costs for patients who need to constantly visit a healthcare facility. There is evidence that 33% of patients, especially elderly individuals, want to have access to remote patient monitoring technologies. Furthermore, 40% of those elderly patients want technologies that notify their caregivers whenever there is an emergency situation [3].

Social network services can be integrated into remote patient monitoring systems [4-12]. This integration enables users to manage their relationships. User-defined relationships are used to disseminate notifications. In emergency situations, notifications lead to quicker interventions and may reduce hospital admission rate.

This paper presents an architecture for notification emission based on integrating social network services to a remote patient monitoring system. As a proof of concept a module was integrated into a remote patient monitoring platform. A mobile application to manage relationships and receive notifications was also developed.

Methods

Our approach to enable notification dissemination in a remote patient monitoring system consists of the integration of social network services. This research evolved over three phases leading to a proof of concept: conceptual phase, design phase, and application development.

Conceptual Phase

Remote patient monitoring systems are specializations of context-aware systems. In this case, context refers to the patients' health status. A context-aware system architecture, as proposed by [13, 14] is shown in Figure 1.

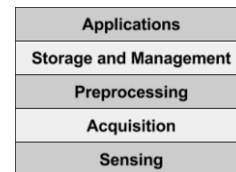


Figure 1 – Context-aware systems architecture, adapted from [13,14]

Context is acquired (Acquisition Layer) using sensors (Sensing Layer); then represented (Preprocessing Layer), persisted and made available (Storage and Management Layer) to the applications residing on the top layer.

Social networks model and manage relationship contextual information [15]. As a context-aware system, a remote patient monitoring system may use the relationship between entities to provide contextual information [16].

Design Phase

Following the pattern present in social network systems, such as Facebook and Twitter, a user receives notifications regarding his/her relationships. A relationship management service can be seen as an event notification service in a publish/subscribe architecture [17]. Creation and destruction of relationships are mapped to subscribe and unsubscribe operations as shown in Table 1.

Subscriptions are content-based [18]. Messages sent to the event notification service have a field specifying the user who subscribed to the service. Information in this field is used to forward notifications.

Table 1 – Relationship manager correspondence with event notification service

	Relationship manager	Event notification
Operation	Create relationship	Subscribe
	Destroy relationship	Unsubscribe

Figure 2 shows a message example formatted using JSON.

```
{
  "id": 4,
  "payload": "message"
}
```

Figure 2 – Message formatted using JSON

Remote patient monitoring systems following this approach are in accordance with the steps proposed by Morrissey [3]; particularly when it comes to notifying designated responders, initiating rapid interventions.

Results

Architecture

Our proof of concept aims to deliver notifications to the caregivers. Notifications are sent whenever the collected physiological data presents an anomaly. Patients' relationships define the notification recipients.

Figure 3 outlines the architecture. Physiological data are collected and transmitted to the UbiCare platform (1, 2 and 3). Users can also interact with the platform using other applications that may be developed (3).

When a user accesses the UbiCare Social application, a device identifier is requested (A) and set (B). This device identifier is maintained in the UbiCare platform (C). To distribute a notification, the UbiCare platform sends a message and a device identifier list to the push service (D). Device identifiers are processed by the push service and the messages are sent to the caregiver and family devices.

Scenario

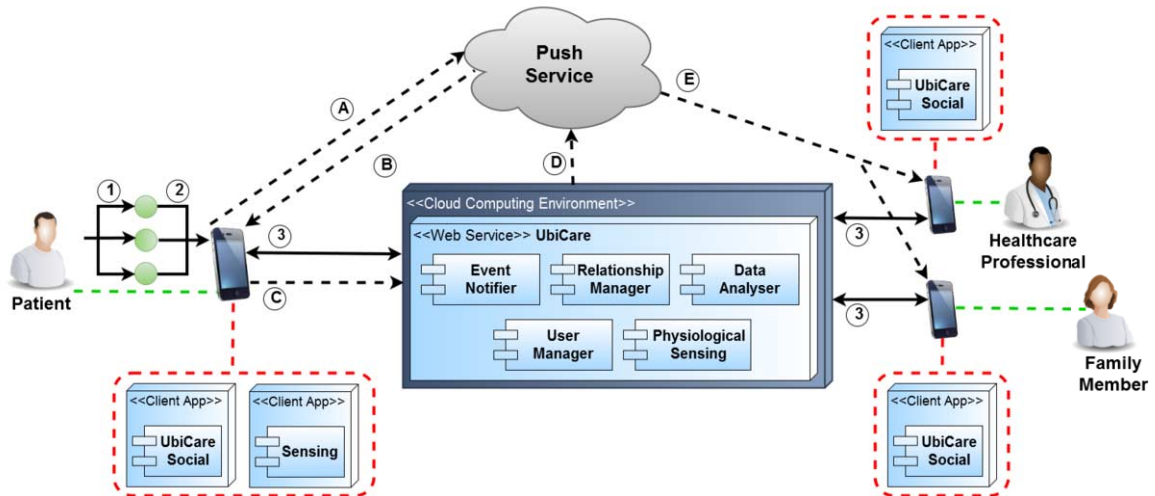


Figure 3 – Architecture

Application Development

Two applications were developed: a relationship manager and a mobile application.

The Relationship Manager is a module integrated into UbiCare, a remote patient monitoring system being developed at Instituto de Informática [19]. This module was built using Flask¹, a web development framework for Python. Relationship data are stored in a MySQL database.

The mobile application, UbiCare Social, consumes the services provided by the Relationship Manager, allowing users to manage their relationships and to receive notifications. This application runs on Android operating system. In addition, Google Cloud Messaging (GCM) push service was used to distribute notifications.

Consider a scenario with three actors: one patient, one healthcare professional and the patient's relative. Their relationships are configured as shown in Figure 4.



Figure 4 – Relationship instance

In a situation in which the patient is an elderly person in post-operative care, remote patient monitoring systems allow the patient to be transferred to home, reducing the number of

¹ <http://flask.pocoo.org/>

inpatients. Healthcare professionals and family members can assist the patient at home, with minimum impact on the patient's privacy, as they will be notified whenever a physiological anomaly is detected.

UbiCare Social

Relationship Manager services are accessed via an Android application, named UbiCare Social. Considering the Healthcare Professional from Figure 4, UbiCare Social behaviour is as shown in Figures 5(a), 5(b), 5(c) and 5(d).

Whenever an anomaly is detected, a message with the Patient identifier is sent to the Event Notifier. Carers' device identifiers are retrieved and the push service (GCM) is triggered. Healthcare Professional and Family Member will receive a notification as the one show in Figure 6.

Discussion

There are research efforts to distribute healthcare related notifications using social network systems [4-12]. Facebook, Twitter, and Google Plus were studied in order to check their suitability to solve this problem [6, 8]. Popular social network platforms have big user bases and provide services via APIs. Despite this, users have minimum control over their data and the platform functionalities. Furthermore, users tend to lack confidence in popular social network platforms.

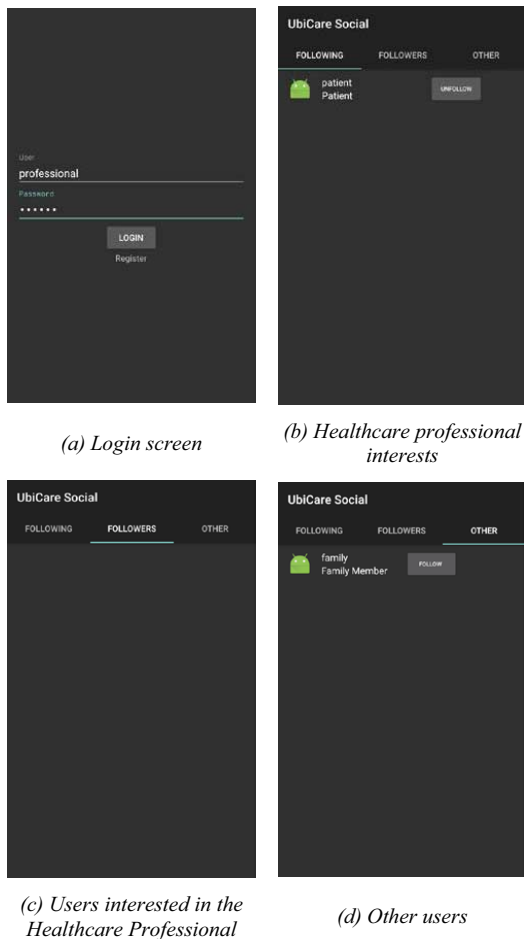


Figure 5 – Mobile application behaviour

Dedicated social networks, such as the one presented in this paper, have the potential to mitigate data ownership, functionalities extension, and trust issues; especially in a research scenario.

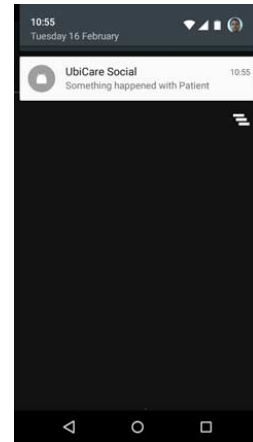


Figure 6 – Notification issued by the UbiCare platform

Conclusion

Social network services may enhance remote patient monitoring systems. This integration leads to quicker interventions and may reduce hospital readmission rate.

Further research can be undertaken in the following areas: user acceptance testing, security, privacy, conformance with norms and standards, inclusion of social media services such as an audio-visual communication channel.

Acknowledgements

The research for this paper was financially supported by CAPES (Coordination for the Improvement of Higher Education Personnel).

References

- [1] World Health Organization, *Global health workforce shortage to reach 12.9 million in coming decades*, WHO (2013), Available at: <http://www.who.int/mediacentre/news/releases/2013/health-workforceshortage/en/> Last access: 12/19/2016.
- [2] World Health Organization, *World report on ageing and health*, WHO (2013).
- [3] J. Morrissey, *Remote patient monitoring: How mobile devices will curb chronic conditions: Factor in the time and cost of implementing data-driven technology*, Medical Economics (2014), Available at: <http://medicaleconomics.modernmedicine.com/medical-economics/content/tags/chronic-care/remote-patient-monitoring-how-mobile-devices-will-curb-c?page=full> Last access: 12/19/2016.
- [4] A. Bhowmik, F. Khendek, M. Hormati, and R. Gliotho, *An architecture for m2m enabled social networks*, In 14th Annual Mediterranean Ad Hoc Networking Workshop (2015), 1–8.
- [5] A. Triantafyllidis, V. Koutkias, I. Chouvarda, and N. Maglaveras, *A pervasive health system integrating patient monitoring, status logging, and social sharing*, IEEE Journal of Biomedical and Health Informatics (2013), 17(1):30–37.
- [6] C. Khorakhun, and S. Bhatti, *Alerts for remote health monitoring using online social media platforms*, In IEEE 15th International Conference on e-Health Networking, Applications Services (2013), 177–181.
- [7] C. Khorakhun, and S. Bhatti, *Remote health monitoring using online social media systems*, In 6th Joint IFIP Wireless and Mobile Networking Conference (2013), pages 1–8.
- [8] C. Khorakhun, and S. Bhatti, *Using online social media platforms for ubiquitous, personal health monitoring*, In IEEE 16th International Con-

- ference on e-Health Networking, Applications and Services (2014), 287–292.
- [9] C. Khorakhun, and S. Bhatti, *Wellbeing as a proxy for a mhealth study*, In IEEE International Conference on Bioinformatics and Biomedicine (2014), 32–39.
 - [10] C. Norval, J. Arnott, N. Hine, and V. Hanson, *Purposeful social media as support platform: Communication frameworks for older adults requiring care*, In 5th International Conference on Pervasive Computing Technologies for Healthcare (2011), 492–494.
 - [11] M. Domingo, *A context-aware service architecture for the integration of body sensor networks and social networks through the IP multimedia subsystem*, IEEE Communications Magazine (2011), 49(1):102–108.
 - [12] Y. B. D. Trinugroho, F. Reichert, and R. Fensli, *eHealth Smart Home Environment Service Platform - Enabling Remote Monitoring and Service Composition through Social Media*, In Proceedings of 5th International Conference on Health Informatics (2012), 434–438.
 - [13] M. Baldauf, S. Dustdar, and F. Rosenberg, *A survey on context-aware systems*, International Journal of Ad Hoc and Ubiquitous Computing (2007), 2(4):263–277.
 - [14] G. W. Musumba, and H. O. Nyongesa, *Context awareness in mobile computing: A review*, International Journal of Machine Learning and Applications (2013), 2(1).
 - [15] A. M. Kaplan, and M. Haenlein, *Users of the world, unite! The challenges and opportunities of social media*, Business Horizons (2010), 53(1):59–68.
 - [16] A. Zimmermann, A. Lorenz, and R. Oppermann, *An operational definition of context*, In Modeling and using context, Springer (2007), 558–571.
 - [17] V. J. Setty, *Publish/Subscribe for Large-Scale Social Interaction: Design, Analysis and Resource Provisioning*, PhD Thesis, 2015.
 - [18] P. T. Eugster, P. A. Felber, R. Guerraoui, and A.-M. Kermarrec, *The many faces of publish/subscribe*. ACM Computing Surveys (2003), 35(2):114–131.
 - [19] E. Germano, D. Battisti, H. A. Ribeiro, and S. T. Carvalho, *Ubiquitous Care Plan for Monitoring Patients at Home* (in Portuguese). In Congresso Brasileiro de Informática em Saúde (2016), 849–858.

Address for correspondence

Hugo A. Ribeiro
 hugoalmeida@inf.ufg.br

Trust Model for Protection of Personal Health Data in a Global Environment

Pekka Ruotsalainen^a, Bernd Blobel^b

^a School for Information Science, University of Tampere, Finland

^b Medical Faculty University of Regensburg, Germany

Abstract

Successful health care, eHealth, digital health, and personal health systems increasingly take place in cross-jurisdictional, dynamic and risk-encumbered information space. They require rich amount of personal health information (PHI). Trust is and will be the cornerstone and prerequisite for successful health services. In global environments, trust cannot be expected as granted. In this paper, health service in the global environment is perceived as a meta-system, and a trust management model is developed to support it. The predefined trusting belief currently used in health care is not transferable to global environments. In the authors' model, the level of trust is dynamically calculated from measurable attributes. These attributes describe trust features of the service provider and its environment. The calculated trust value or profile can be used in defining the risk service user has to accept when disclosing PHI, and in definition of additional privacy and security safeguards before disclosing PHI and/or using services.

Keywords:

Trust; Health Records, Personal ; Privacy

Introduction

Starting from Hippocratic time, trust has been one of the main cornerstones in successful healthcare. Until now, trust and distrust have not been big questions in today's health care. Instead, it is expected that both health care services and their information systems occur in a controlled environment where ethical codes and laws guarantee fair and trustworthy information processing. It is also expected that patients intrinsically trust the health care service provider, and believe that information systems and networks that communicate, process, and store patients' personal health information (PHI) are trustworthy. In other words it is argued that predefined organizational trust is sufficient, and security based access controls guarantee privacy [1].

However, health care in general is in transition. Health care services and information systems are increasingly provided cross-organizationally, across boundaries, and cross jurisdictionally. New services models such as digital health, personal health, health eco-systems, and ubiquitous health take place in global information networks. They are built on using modern information and communication technology (ICT), global communication networks, and applications as service. This implies that service providers, customers, patients, the PHI and applications can operate in different contexts and jurisdictions. Furthermore, the PHI is increasingly collected,

used, communicated and stored in environments not regulated by health care or privacy laws, global guidelines, codes of conducts; and fair information processing rules are implemented just poorly or even worse, not at all. It is also common that service provider, service user (a person or patient), medical practitioners, secondary users of PHI and health software developers can all have their own notion of how PHI should and can be used and protected. This may basically differ from the expectations of the data subject and his or her local regulations.

From the standpoint of privacy and trust, the modern global and distributed health service environment is challenging. First, in the context of cross-jurisdictional e-health and personal health services, contextual trust and privacy features cannot be predicted or measured in advance. Secondly, the network itself is insecure, and the data collector or customer has few or no tools to measure the level of trust. Furthermore, he/she has limited or no power to enforce informed privacy and security decisions concerning the trustworthiness of services and control how, by whom and for what purposes PHI is collected and processed [2].

It is widely accepted that trust and privacy are key enablers for global health care and the use of personal health services. Using the self-regulation principle, industry has developed trust, security and privacy rules for eCommerce. Unfortunately, those rules are most of all developed to support industry's own business needs, expecting that customers blindly trust on the service provider and accept rules as they are (i.e. take-or-leave principle). Yuan and Ruotsalainen et al. have mentioned that increasingly modern health care and ubiquitous health services are dynamic and take place in insecure and uncertain environment where no predefined trust cannot be expected [2, 3]. This indicates that current rules used by eCommerce cannot be moved to health services as such.

The authors state that health information is highly sensitive requiring special protection. To enable trustworthiness of global health care and the use of personal health services, there is an urgent need for practical and easy-to-use solutions for trust measurement, trust creation and management. Without such prerequisites, it can be dangerous for a service provider to disclose PHI, and for a customer (persons and patients) to use offered services. In global environments, it is also necessary that customers using services and service providers disclosing the PHI can make rational and information based choices concerning additional safeguards needed, in advance.

This paper is based on the following assumptions: Privacy and trust are interrelated concepts in a way that less trust requires more privacy protection. Trust is situational and context-

dependent. Knowing the trust level of a health service provider enables the service user (patient, person or organization) to make rational choices concerning to what extent it has the willingness to use services, and what amount of PHI it is ready to disclose at certain level of trustworthiness. In global environments, trust features expressed in the form of trust value or trust profile enable the service user to define necessary safeguards before starting the use of services.

Previous research

Trust is a multifaceted, context-dependent concept, and a term with many meanings. There is no globally agreed definition for it. Widely used trust models are trusting belief, organization/institutional trust, dispositional trust, recommended trust, direct trust, and computational trust [4]. In the context of global health services, trust can be seen as a process of practical reasoning that leads to the decision to interact with somebody [5]. Institution-based trust deals with structures (e.g. legal protections) that make an environment trustworthy. Institutional trust is the belief that needed structural conditions are present [6]. System trust represents the extent to which a customer believes that the proper structures are in place, i.e. that reasonable safeguards are in place to reduce risk. These safeguards may be represented in form of regulations, guarantees, or stabilizing intermediaries [7].

Ruotsalainen et al. have noted that in networked and ubiquitous health service systems dispositional trust, recommended trust, and direct trust are not much stronger than belief, and organizational trust is static [2]. Because the use of global health services is increasingly dynamic, it requires the possibility to make online trust decisions.

Trust models are often based on use of a pure numerical approach. The mechanisms used to calculate trust values range from simple aggregation of values to the use of probability theory, fuzzy logic, or the use of entropy [8]. The number of past experiences, observation interaction, and recording are also widely used [5].

According to Saadi et al., previously discussed “classical” trust approaches that cannot be adapted to networked and ubiquitous environments such as cross-jurisdictional healthcare and personal health systems where the unpredictability and unreliability of service location, contextual features, regulation and rules make mechanisms inappropriate [9]. Viljanen et al. have defined a trust formulation process using trustors contextual attributes and actions, information attributes, social and ethical attributes and third party information (e.g. certificates or recommendations) [10].

Trust models are developed especially for multi-agent systems, Mobile Area Networks (MANET), and open dynamic and ubiquitous environments [5,11,12]. In MANETs, trust is typically evaluated using transaction history of past interactions and transactions and others recommendations. Another approach is the use of trusted third parties and certificates [13]. Hereby, calculated trust values are deployed as estimates of the service provider’s trustworthiness [14].

Ruotsalainen et al. have proposed the following attributes for trust calculation in pervasive health: trustor’s environmental factors and contextual features, ICT systems properties, privacy

policy, predictability, transparency and openness, and system’s regulatory compliance [2].

Methods

In this paper, the collection and use of the PHI in global environment is perceived as a meta-system characterized by its structure, functions, behaviour, and relevant stakeholders. Using system modelling methods, system analysis and system engineering techniques, a conceptual trust management model for the protection of PHI in global environment is developed.

Based on a careful analysis of findings and proposals from research published in journals and conference proceedings, measurable trust attributes are identified. Attributes are aimed for the evaluation of the level of trust of different kind of health providers and secondary users of the PHI in existing in global environment.

Results

Because the global health service environment forms a meta-system, trust should be created between its actors. According to Saadi, Sabater-Mir and Zheng [2, 8, 9, 14], the authors state that belief and recommendation based trust solutions are too weak. Instead, dynamic system trust that is based on service provider’s real life measurable features is the most promising approach.

Trust Management Model

Because the use of PHI takes place in different contexts, contextual trust approach is needed. According to Jøsang et al., contextual trust describes the extent the data subject can expect that necessary services and institutions are in place in order to support trustworthy communication; and trust implies a decision [15]. As discussed earlier, belief as well as dispositional and recommended trust approaches cannot be used in dynamic, distributed environments, and meaningful trust decisions are impossible without reliable information. Therefore, the approach of calculated contextual system trust that is based on measured features of service providers is selected for the model. Thereby, policies can be used to define what is permitted or prohibited, and what security and privacy obligations the service provider must perform in advance.

The proposed model for trust management in global health services is shown in Figure 1. The model is focused on the processing of PHI in different contexts and environments. The model is developed and presented using UML.

In the model, service providers can be either regulated or nonregulated health service providers, or other entities processing PHI (e.g. secondary users). Non-regulated service providers include institutions beyond regulated healthcare establishments such as personal health systems, personal health and ubiquitous-health services. From a data processing perspective, service providers are represented by different instances such as data creator, data controller and data processors [16]. The data controller manages, stores, and discloses personal health information to data processors. Data processors use disclosed health data on behalf of data controllers [16].

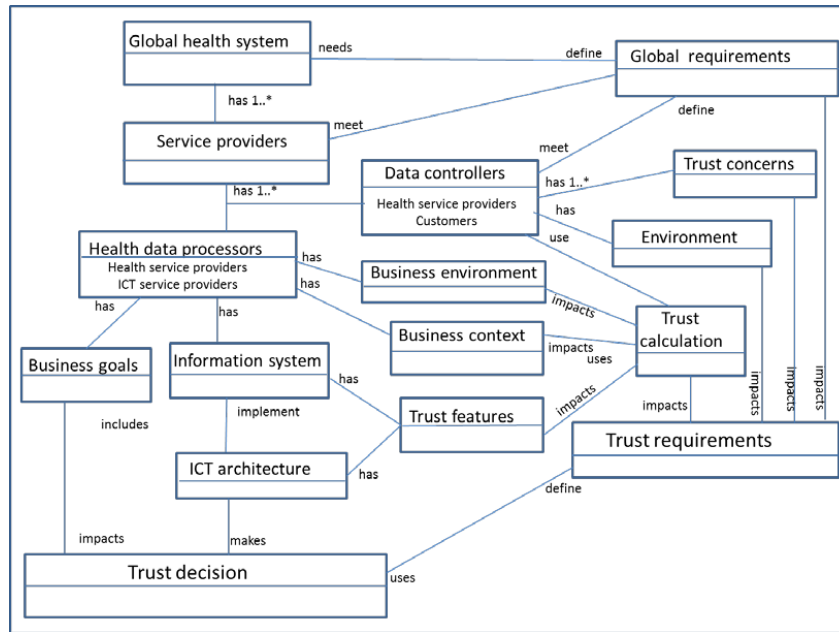


Figure 1 - A conceptual model for the trust management in global health

The data controller can be a health care provider/providing organization or the data subject (a person or patient). The data processor is any entity deploying received PHI during its service process. The data processor has own business goals such as offering health service or using PHI for research. In global health settings, the data controller often has no predefined trust to the service provider to whom PHI is disclosed to. Instead it has trust concerns.

The data processor deploys an information system with an appropriate ICT architecture (the concept of a system covers both components and processes, i.e. structure and behaviour of that system). One component of the ICT system is the trust decision application. It is typically an Artificial Intelligence solution.

The environment includes national and international regulations, laws and norms guiding the processing of PHI, e.g., security and privacy regulations.

The business environment is a combination of external and internal factors such as organizational rules and constraints in the framework of national and international regulations. The business context of the data processor in the system in question includes expectations of other parties involved such as the data controller/customer, but also process-specific constraints, technologies, etc.

Global requirements include ethical principles and codes of conduct, international Golden Rules (e.g. Fair Information Practice Principles, OECD principles), standards and international certification requirements (e.g. possible future global privacy regulations for health services).

Because trust and privacy are interrelated in such a way that lower trust requires more privacy safeguards, both trust requirements (e.g. audit-log for transparency, policy based access control as properties of the ICT system, standards for reliability) and privacy requirements such as anonymization and denying the post-release of the PHI have to be managed.

Trust creation provided by the data controller requires that the level of trust of the data processor can be defined, formalized and communicated. In the model, the Trust calculator service (e.g. an agent application) collects necessary information for the calculation of trust level or trust profile of the service provider and communicates this information.

The Trust Requirements service uses information received from the Trust calculator, Trust concerns, Global requirements and service providers Trust features to define rules (policies) the service provider offering health service should follow. System requirements such as audit-log, policy based PHI management, notification, and privacy requirements such as anonymization and denying the post-release of the PHI are expressed in the form of computer understandable policies as defined in ISO 22600. In the proposed model, each service provider is first authenticated and then assigned to a trust value or trust profile. Trust and privacy requirements associated to a service provider are processed by its trust decision application.

The presented model is suitable for both static and dynamic online situations in global environment. For example, the Trust calculator can be a Certification Authority (a Trust CA) that generates trust certificates for service providers, and compares them against data controller's certificate.

Trust attributes

In the developed model, trust attributes are needed and used to measure the amount of trust. Researchers have suggested more than 40 different trust attributes [6, 14, 17, 18, 19, 20] (Table 1). Seppanen et al. have defined benevolence, competence, fairness, honesty, moral integrity, motivation, predictability, and reputation as most common trust attributes [21].

The challenge with most of trust attributes shown in Table 1 is that they are difficult to conceptualize and to measure. Therefore, it is also difficult in global environment to generate common understanding for attributes. For overcoming that problem, the concepts should be represented explicitly using ontology representation tools.

Table 1 - Common trust attributes

Ability	Competence	Likeability	Responsibility
Authenticity	Data quality	Moral integrity	Receptivity
Benevolence	Fairness	Motivation	Reciprocity
Certificates	Frankness	Recommendation	Risk
Credibility of promises	Judgment	Openness	assessment
Confidence	Honesty	Past interactions	Security
Consistency	Habitualization	Privacy policy	System
Discreteness	Integrity	fulfilment	reliability
Integrity	Intention	Reputation	Togetherness
Confidentiality	Institutionalization	Predictability	Transparency
	Legal requirements		Tracking
			Transactions
			Trustee's promises

To make attributes acceptable and implementable at global level, the authors propose the following set of attributes for the calculation of trust when health services are used in global environment:

- Ability and willingness
- Integrity
- Openness and transparency
- Properties if service providers manage the ICT system
- Predictability of promises
- Reliability of service provider's promises
- Service provider's environmental factors and contextual features
- Service provider's regulatory compliancy
- Willingness to follow rules (policies) the data processor define

Attributes presented above form a minimum set of attributes that can be measured in real life situations.

Ability can be calculated from direct measurement and/or from the systems' past history [2]. Integrity addresses that the service provider accepts rules and meets its promises [18]. This can be measured using systems' history and other's witnesses. Openness and transparency means that service provider's security and privacy policies, audit trail, standards and laws used, evaluation and risk assessment documents are openly available, and the security and privacy breaches will be notified to the data controller. Properties of the service providers' ICT systems can either be identified from evaluation or assessment reports, with the help of system documents and features expressed in contract documents, or from trust certificates. Predictability concerning service provider's promises can be measured using systems history or continuous monitoring. Service provider's contextual and environmental features can be resolved by available information concerning service provider's location and business goals. Regulatory compliance can be measured using conformance assessment and the regulatory compliance documents. Willingness to follow rules (policies) the data controller has defined can be measured either by direct measurements or by monitoring.

In the model, trust value can be expressed using the scale proposed by Liu [22]: Compromised or malicious, unable to determine trust-level, low trust level, medium, fairly high trust level, and extremely high trust level. For more detailed trust creation, the data controller can use rich trust profiles received from the Trust calculator [2]. Based on calculated trust values, the data subject can define service provider specific data processing policies for all organizations and persons

participating in the service provision chain, and for all secondary users [2].

Discussion

In this paper, the authors proposed a novel trust formulation and management model for health care and health information systems operating in the global information space. The model enables the data controller to disclose PHI, and the customer to use networked health services, by creating and managing contextual trust across geographical, cultural and jurisdictional borders. The authors have also proposed nine measurable trust attributes which the data controller can use in defining additional service provider specific privacy requirements.

The proposed model is flexible. For example, it accepts the use of certificates. Unfortunately, a typical certificate represents only a digital identity of the users, and is static [9]. Therefore, a trust certificate that can be used for the evaluation of trust level requires much richer information such as the trust profile.

In the model presented, the data controller (data subject or organization controlling the use of PHI) can make informed policy decisions by balancing service benefits expected and own privacy needs against trust level of the data processor. A strength of this model is that it enables the calculation of trust level/profile in cases where only incomplete information of data processor's trust features is available. The latter situation generates low trust value. In this way, the proposed solution is proactive and stresses the data processor to support openness and transparency. Challenges include the development of trust calculation services and the global agreements on trust attributes. This might require political and legal actions at global level. It is also necessary to test the feasibility of proposed attributes, and standardize their presentation and meaning. In the future, it is also necessary to demonstrate that the proposed solution is technically valid, reliable and easy to use. Globally, the biggest challenge is to make the principle of direct measurement based on calculated trust accepted by health industry, healthcare professionals and organizations. International political, legal, regulatory and organizational actions are needed to make this true.

Conclusion

The authors have developed a conceptual model for trust management in global, cross-organizational and cross-jurisdictional health service environment. The model enables the data controller (a person or organization) to evaluate the level of trust-worthiness of the data processor before starting to use services or to disclose PHI. Both regulated healthcare and nonregulated health services models are supported. For trust evaluation/calculation, a set of practical and measurable trust attributes is proposed.

In the model, trust level is expressed as trust value or trust vector. The model enables the data controller to define for the service provider minimum privacy and security safeguards required to be qualified trusted.

References

- [1] Ruotsalainen P, Blobel B, Seppälä A, Sorvari H, Nykänen P, A Conceptual Framework and Principles for Trusted Pervasive Health, J Med Internet Res 2012;14(2):e52 doi:10.2196/jmir.1972.

- [2] Ruotsalainen P S, Blobel B, Seppälä A, Nykänen P, Trust Information-Based Privacy Architecture for Ubiquitous Health, *JMIR Mhealth Uhealth* 2013;1(2):e23, doi:10.2196/mhealth.2731.
- [3] Yuan W, Guan D, Lee S, Lee Y-K The Role of Trust in Ubiquitous Health Care, 9th International Conference on e-Health Networking, Application and Services, IEEE Xplore July 2007 DOI: 10.1109/HEALTH.2007.381660.
- [4] Sabater J, Sierra C: Review on computational trust and reputation models, *Artif Intell Rev* (2005) 24: 33. doi:10.1007/s10462-004-0041-5.
- [5] Pinyol I, Sabater-Mir J, Computational trust and reputation models for open multi-agent systems: a review, *Artif Intell Rev* (2013) 40:1–25, DOI 10.1007/s10462-011-9277-z.
- [6] McKnight DHC, Choudhury V, Kacmar C, Developing and Validating Trust Measures for e-Commerce: An Integrative Typology, *Information Systems Research*, Volume 13 Issue 3, September 2002 Pages 334-35, doi:10.1287/isre.13.3.334.8.
- [7] Gray E, O'Connell P, Jensen C, Weber S, Seigneur JM, Yong, Towards a Framework for Assessing Trust-Based Ad mission Control in Collaborative Ad Hoc Applications, 2002 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.4803&rep=rep1&typ>.
- [8] Sabater-Mir J, Towards the next generation of computational trust and reputation models, *Proceedings of the Third International Conference on Modeling Decisions for Artificial Intelligence (MDAI'06)*, 19-21, ISBN:3-540-32780-0 978-3-540-32780-6, doi>10.1007/11681960_4.
- [9] Saadi R, Pierson JM, Brunie L. (Dis)trust Certification Model for Large Access in a Pervasive Environment, *International Journal of Pervasive Computing and Communications*, (2005) Vol. 1 Is: 4, pp.289 - 299 ISSN: 1742-7371.
- [10] Viljanen, Towards an Ontology of Trust, In Katsikas S, Lopez I and Pernul G (Eds.) *Trust, Privacy and Security in Digital Business*, *Proceedings of the 2nd International Conference TrustBus 2005*, Copenhagen, Denmark ,August 22-26 2005, LNCS 3592 Springer.
- [11] Matt PA, Morge M, Toni F, Combining statistics and arguments to compute trust, *AAMAS '10 Proceedings of the 9th International Conference on Autonomous Agents and Multi-agent Systems: volume 1 - Volume 1* Pages 209-216, 2010, ISBN: 978-0-9826571-1-9.
- [12] Velloso P B, Laufer RP, Cunha D, Duarte O, Pujolle G, Trust Management in Mobile Ad Hoc Networks Using a Scalable Maturity-Based Model I, *EEE transactions on network and service management*, VOL. 7, NO. 3, September 2010.
- [13] Omar M, Challal Y, Bouabdallah A Certification-based trust models in mobile ad hoc networks: A survey and taxonomy, *Journal of Network and Computer Applications* October 30, 2011, <https://hal.archives-ouvertes.fr/hal-0064449>, 2011.
- [14] Zheng Y, Holtmanns S, “Trust Modeling and Management: from Social Trust to Digital Trust”, book chapter of *Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions*, IGI Global, 2007.
- [15] Josang A, Ismail R, Boyd V: A Survey of Trust and Reputation Systems for Online Service Provision, *Journal Decision Support Systems*, Volume 43 Issue 2, March, 2007, Pages 618-644.
- [16] EU Data Protection Directive, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, <http://eur-ex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>
- [17] Becerra G, Heard J, Kremer R, Denziger J, Trust Attributes, Methods, and Uses, In *Proceedings of the Workshop on Trust in Agent Societies, AAMAS-2007 Honolulu Hawaii USA*; May 2007, 1-6, https://www.researchgate.net/publication/28814176_Trust_attributes_methods_and_uses.
- [18] Hussin Ab RC, Macaulay L, Keeling K. The importance ranking of trust attributes in e-commerce Website. 2007 Presented at: *Proceedings of the 11th Pacific-Asia Conference on Information Systems*; Jul 3-7, 2007; Auckland, New Zealand URL: <http://www.pacis-net.org/file/2007/1247.pdf>.
- [19] Kim, Dan J. Ferrin, Donald L. RAO, H. Raghav. A Trust-Based Consumer Decision Model in Electronic Commerce: The Role of Trust, Risk, and Their Antecedents. (2008). *Decision Support Systems*. 44, (2), 544-564. Research Collection Lee Kong Chian School of Business.
- [20] Mayer RC, James JH, Schoorman FD, An integrative model of organizational trust, *Academy of Management Review* 1995, Vol. 20. No. 3, 709-734.
- [21] Seppanen R, Blomqvist K, Sundqvist S, Measuring inter-organizational trust – a critical review of the empirical research in 1990-2003, *Industrial Marketing Management* 26(2007), 249-265.
- [22] Liu Z, Robert A W J, Thompson R A, A Dynamic Trust Model for Mobile for Mobile Ad-Hoc Networks, *Proceedings of the 10th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS'04)* 0-7695-2118-5/04, 2004 IEEE.

Address for correspondence

Pekka Ruotsalainen, Adjunct Profesor, Research Professor emeritus, University of Tampere, Finland, Chair IMIA WG SiHiS. E-mail: pekka.ruotsalainen@uta.fi.

PubMedReco: A Real-Time Recommender System for PubMed Citations

Hamman W. Samuel, Osmar R. Zaiane

Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada

Abstract

We present a recommender system, PubMedReco, for real-time suggestions of medical articles from PubMed, a database of over 23 million medical citations. PubMedReco can recommend medical article citations while users are conversing in a synchronous communication environment such as a chat room. Normally, users would have to leave their chat interface to open a new web browser window, and formulate an appropriate search query to retrieve relevant results. PubMedReco automatically generates the search query and shows relevant citations within the same integrated user interface. PubMedReco analyzes relevant keywords associated with the conversation and uses them to search for relevant citations using the PubMed E-utilities programming interface. Our contributions include improvements to the user experience for searching PubMed from within health forums and chat rooms, and a machine learning model for identifying relevant keywords. We demonstrate the feasibility of PubMedReco using BMJ's Doc2Doc forum discussions.

Keywords:

PubMed; Medical Informatics Applications; Information Storage and Retrieval

Introduction

PubMed is the de facto tool for searching biomedical and life sciences literature. PubMed is comprised of the MEDLINE (Medical Literature Analysis and Retrieval System Online) bibliographic database, which covers academic journals on medicine, pharmacy, nursing, dentistry, and medical care, among others, with over 23 million trustworthy and peer-reviewed articles[1]. Essentially, PubMed is the web interface to the MEDLINE database. PubMed is typically used from its home page of the National Center for Biotechnology Information (NCBI) website. Based on the keywords entered in the search engine, matching citations are returned, which can be further explored by clicking on each citation to view abstracts or full-text.

The MEDLINE database is indexed using the Medical Subject Headings (MeSH), an indexing medical vocabulary[2]. To facilitate natural language usage in the PubMed search engine, PubMed uses the process of Automatic Term Mapping, which matches non-MeSH keywords to the MeSH index. A MeSH translation table is used for each query issued by the user to map the natural language keywords to equivalent MeSH keywords. In order to broaden search results, PubMed also leverages mappings of the search query keywords derived from the Unified Medical Language System (UMLS).

UMLS is a collection of biomedical vocabularies and standards, and PubMed can leverage these vocabularies to expand search queries using hypernymns, hyponyms, synonyms, and other

semantic relationships[3]. The interactions between PubMed, MeSH, UMLS, and MEDLINE are depicted in Figure 1.

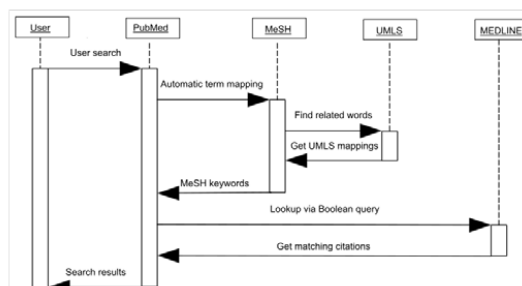


Figure 1 – Sequence diagram depicting interactions between PubMed, MeSH, UMLS and MEDLINE during search

However, using the PubMed web interface for searching is not ideal when users are in a contained environment such as a chat room or a forum discussion thread. Users would have to leave their current web page in order to go to the PubMed website and retrieve the information they need to look up. Furthermore, users need to have a sense of the context of the overall conversation and the key topics being discussed.

Our proposed recommender system integrates PubMed citations into a unified user experience so that medical citations relevant to the on-going conversation are conveniently accessible for the users. From our survey of existing literature, our proposed system is the first of its kind in the medical domain. It should be noted that access to reading an article depends on the individual user's subscription to PubMed; our system displays only the citation, and a clickable hyperlink to the article conveniently within a chat environment interface.

A recommender system is a set of software tools and algorithms that can give useful suggestions to users[4]. The suggestions are given within the context of the user's domain of interest such as what items to buy or shop for, which new people to connect with, or which new movies to watch. Recommender systems are useful from both the perspectives of the content producers (such as sellers, blog writers, videographers, movie makers, and so on) and the content consumers (such as buyers, readers, fans, etc.). Recommendations enable sifting through large amounts of information previously too massive or complicated to practically navigate. Recommendations also enable users to focus on things that interest them personally, thereby increasing the findability of information.

There are various methods for generating recommendations: content-based, collaborative, community-based, demographic, knowledge-based[4]. Content-based recommendations use keywords to suggest new items that are historically similar to

previous items that a user may have liked or bought in the past. In collaborative recommender systems, suggestions for one user are based on what other users with similar profiles have liked. Community-based recommendations leverage the preferences of users' friends, a popular notion in Social Networking Sites (SNS). The demographic approach uses age, gender, ethnicity, and other demographic information about a user, as well as the items they like or buy. These properties are matched to demographic stereotypes such as English-speaking customers being directed to US-based websites. The knowledge-based approach is to recommend items using specific domain knowledge about how certain item features meet user needs and preferences.

Recommendation systems in chat rooms have been researched in other domains. PALTASK is a personalized automated context-aware web resources listing tool that suggests resources of common interests to users in a chat room[5]. These resources include videos, web articles, and social media links relevant to the conversation. PALTASK performs contextual analysis, and determines the context of the conversation based on each conversing user's profile, called Personal Contextual Sphere (PCS), and also from the chat conversation contents. The PCS is determined when the user mentions specific keywords related to their personal interests. As an example, context analysis can determine that a user mentioned a music video in the chat. PALTASK can then display links to that music video within the chat window interface.

Another real-time chat recommender system has been proposed for e-commerce websites[6]. The recommender system is based on profiling users while they chat. A user may be interacting with another user to chat about buying things, or may be chatting with a seller. The profile contains five parameters: "unusual", "cute", "cool", "simple", and "luxurious", with each parameter having a value 1-5, with 5 being the strongest value. The recommender updates this profile by analyzing chat conversations based on inferring positive and negative connotations of words in the chat with the parameters, and using words with associated positive feedback to query the product catalog and recommend items.

Another research work of interest looks at the role of temporal dynamics in product recommendations[7]. The research focuses on evolving user preferences over time, arguing that traditional methods do not take into account the temporal changes of user preferences. For example, if a user bought a particular baby food from an e-commerce website, it is very likely that they would not be interested in seeing recommendations related to their purchase after a while because the dietary needs of a child would be changing as the child grows. To recommend new items, collaborative filtering-style similarity metrics are used based on transactional history. By using these metrics in cluster-based and graph-based modeling, the system retrieves similar items to recommend.

Stream and event processing systems have also been explored regarding real-time recommendations[8,9]. Streams are sequences of events, and stream systems require high throughput processing, making them ideal for the real-time recommendation approach.

Social news is an example of stream systems, where new stories are continuously being added, and user ratings can instantly affect recommendations for related interesting news. The recommendation engine, StreamRec, uses collaborative filtering and works in two phases: model building and recommendation generation. In the model building phase, a similarity score is computed for user-rating pairs. In the recommendation generation phase, items are given a predicted rating based on the similarity scores, and items with a high predicted rating are suggested.

In both PALTASK and the e-commerce recommender, incorporation of temporal context has not been considered. For instance, if keywords are extracted from the entire conversation, some keywords may be outdated because they were discussed hours ago. Moreover, even if the latest chat conversation is being used, it might not give the best recommendations. The latest message might diverge significantly from the overall conversation, and could be an outlier. For the StreamRec engine and the products recommender, reliance on a history of ratings could potentially result in the cold start problem, whereby the initial set of ratings have to be determined arbitrarily in the absence of a history of ratings[10]. Another issue with using historical ratings for determining recommendations is sparsity of data, whereby only a few users might choose to give feedback or ratings.

It should also be noted that PubMedReco is a content-based recommender system for temporal conversations, and not a Time-Aware Recommender System (TARS)[11]. The focus of TARS is on temporal relationships between recommendations, while PubMedReco looks at temporal and contextual links between the keywords needed to provide recommendations.

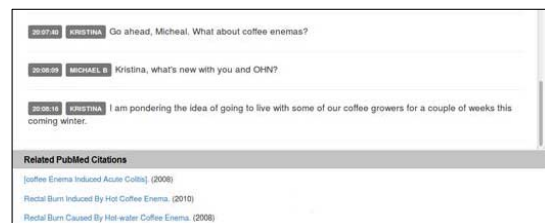
Methods

To demonstrate the feasibility of PubMedReco, we develop a prototype that is populated with conversations taken from a health forum, thereby simulating a chat-like environment where new messages are being added over time. For evaluation, we look at three aspects of accuracy: selection of relevant keywords, agreement about what constitutes keyword relevance, and correlation between recommended citations and selected keywords.

An overview of the prototype is shown in Figure 2, where users are chatting on the topic of coffee enema. For each new message, medical keywords are extracted. A subset of all the keywords extracted are then used to query PubMed and get related citations. In this way, users can view citations related to the overall conversation without needing to go to the PubMed website, or having to determine which specific keywords to use for querying PubMed.

System Design

The PubMedReco system is outlined in Figure 3. As a new message arrives, it is tokenized via regular expressions into individual keywords. Next, common English stopwords are discarded, and the remaining keywords are retained along with relevant information such as the timestamp of the message containing the keyword. In the next step, the incoming keywords from the new message are looked up in an index of all retrieved keywords, which is initially empty. If the incoming keyword exists, only its related information is updated, including the latest message number containing the keyword. Otherwise, the new keyword is added to the index.



(a) Chat interface at time t_1

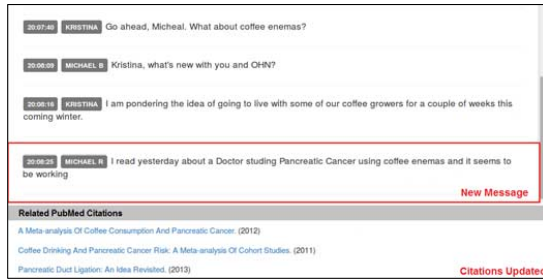
(b) Chat interface at time t_{i+1}

Figure 2 – PubMedReco Prototype

```

1. keywordsIndex = NULL
2. FOREACH newMessage IN allMessages
3.   keywords = RegExTokenizer(newMessage)
4.   keywords = StopWordsFilter(keywords)
5.   FOREACH keyword IN keywords
6.     IF keyword IN keywordsIndex
7.       keywordsIndex.update(keyword)
8.     ELSE
9.       keywordsIndex.add(keyword)
10.  relevantKeywords = NeuralNet(keywordsIndex)
11.  citations = EUtilities(relevantKeywords)
12.  RETURN citations

```

Figure 3 – PubMedReco System Design

The keywords index is then fed to an artificial neural network that selects keywords that are both temporally and contextually relevant. Artificial neural networks or neural nets, are computational models that are based on the structure of the biological brain, where a large collection of neurons and axons can enforce or inhibit signals, and in turn activate different states[12]. Neural nets have been used for modeling binary classification problems, where elements of a given set need to be categorized into two groups based on specific rules. We model the task of determining relevant keywords by inputting various features for each keyword and then outputting 0 if the keyword is irrelevant, or 1 if it is relevant. The keyword features are listed in Table 1.

Neural Net Design

The keyword features are converted to binary form for input into the artificial neural net's neurons, as neuron states are more readily represented as 0 or 1.

This conversion includes the keywords themselves, which are converted to word embeddings using the Word2Vec deep neural network model trained with skip-grams on the entire Doc2Doc dataset[16]. The word embeddings enable each unique keyword to be assigned a corresponding binary vector in the space. Also, keywords with common contexts such as synonyms are positioned close each other in the vector space, and have similar binary vector values.

Table 1 – Neural Net Keyword Features

Property	Description
keyword	Word embeddings representation of keyword
isMed1	Is keyword in the Merriam-Webster's Medical Dictionary ¹
isMed2	Is keyword in SNOMED database ²

¹ Merriam-Webster's Medical Dictionary with Audio API available at <http://www.dictionaryapi.com>

² U.S. National Library of Medicine (NLM)'s Systematized Nomenclature of Medicine (SNOMED), International edition

firstPos	Message number/reference where keyword first appears
lastPos	Message number/reference where keyword last appeared
firstTime	Seconds passed since epoch till first occurrence of keyword
lastTime	Seconds passed since epoch till last occurrence of keyword
frequency	Number of times the keyword has appeared in the chat
numMsgs	Number of messages in the entire conversation

Training Dataset

Initially, the neural net needs to learn how to associate these features to the groupings by using a training dataset with keywords already classified as relevant or irrelevant. The neural net model is trained by manual annotation of a subset of BMJ's Doc2Doc forum discussions dataset³. The annotation process involves manually inspecting the keywords for each new incoming message, and marking their relevancy based on the current context. The Doc2Doc forums allow doctors to have online discussions with other doctors on various health-related topics. The temporal nature of forum conversations is ideal for testing PubMedReco, as forum discussions progress over time like online chats. Also, the technical nature of doctors' conversations makes the dataset suitable for querying PubMed. It should be noted that a forum discussion or chat containing n messages yields n training sets because annotations are made for each new incoming message. The trained model can then be used to predict the relevance of new forum discussions or chats that do not have any manual annotations.

Citations Retrieval

Once the relevant keywords are selected, they are then used to query the PubMed database programmatically using Entrez Programming Utilities (E-utilities), a RESTful programming interface[3]. E-utilities accept natural language queries and converts them into Boolean queries by inserting Boolean operators and using the words as operands. Stemming and lemmatization are also performed on keywords by the E-utilities API which can infer synonyms and other relationships to the query words via UMLS.

E-utilities has options for specifying what citation fields to search within, such as title, abstract, full text (where available), author and others. Our proposed system restricts search to the citation title because our recommendations are ultimately presented as full citation titles. Consequently, the user would decide initial interest or disinterest in the recommendation based on the displayed title. The citations returned can also be sorted using various options available in E-utilities, and we sort by relevancy, which takes into account the frequency of matched keywords within the title. Hence, citations containing more of the search keywords would be ranked higher.

Evaluation Criteria

As mentioned before, three aspects of PubMedReco need to be evaluated for accuracy: neural net, annotations, and recommended citations.

The neural net needs to be appraised for accuracy, in order to determine how it would perform when given datasets that have no annotations. The evaluation is done via the standard precision metric. A sampling out of the total number of training sets generated is selected, and the precision measured. This process is iteratively done in order to see which random sampling provides the best precision value.

The annotations are related to the performance of the neural net. For evaluating the quality of annotations about the relevance of keywords, we use the Kappa score to compare multiple

³ This dataset is no longer available on the BMJ website, but is still accessible via the Internet Archive's Wayback Machine project at <http://web.archive.org/web/20160615110024/doc2doc.bmj.com>

annotations made on the same dataset. As baseline, we use arbitrary annotations so that there is no planned correlation between a keyword and its relevance.

Finally, for evaluating the quality of the recommended citations, we use Normalized Discounted Cumulative Gain (NDCG)[13] to quantify whether the recommended citations indeed contain the keywords that were used to query PubMed. NDCG measures both the number of matching keywords in a search hit, as well as the usefulness of the hit based on its position in the results. As baseline, we use the “Title” sorting option within E-utilities, so that the top- n citations are sorted alphabetically. This sorting option will not rank citations based on number of keyword hits, but rather based on the title’s alphabetic ordering.

Results

Firstly, the results of evaluating the neural net are presented. The Doc2Doc dataset contained a total of 1,400 discussions. Out of these discussions, 10 were used as the training dataset with varying numbers of message threads, averaging 9.80 threads per discussion. The total number of training datasets generated from the 10 discussions was 98. The neural net was trained by iteratively and randomly selecting 50 samples out of the 98 training datasets and choosing the model with the highest accuracy. Figure 4 shows the precision for 20 iterations of the sampling. The average precision was 55.87%, while the highest precision was 60.80%.

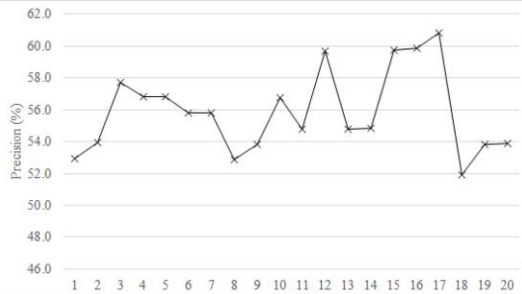


Figure 4 – Precision of Neural Net

Secondly, we present the evaluation of the manual annotations. We selected 10 discussions with annotations, and re-annotated them without cross-referencing the previous annotations. An average Kappa score of 51.87% was achieved, showing borderline acceptable agreement with the annotation method.

Figure 5 shows the Kappa scores for the 10 annotated discussions (M), their counter-part annotations (N), and baseline arbitrary annotations (Base).

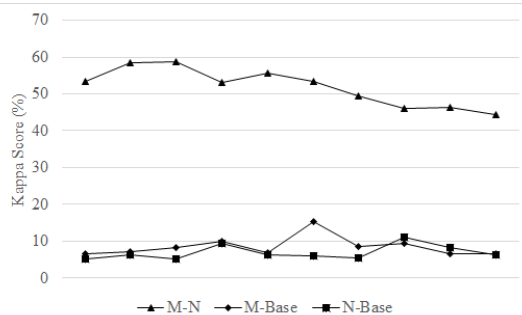


Figure 5 – Kappa Score for Annotations

Thirdly, we present statistics on the quality of the citations using NDCG. We arbitrarily selected 5 discussions and computed NDCG for each query to E-utilities, resulting in 32 data points, and an average number of 6.4 messages in the selected discussions. Figure 6 shows the NDCG values with E-utilities optimal sorting (Relevancy), which averaged 0.798, and also the baseline using E-utilities alphabetic ordering of the citation (Title). Similar results were also achieved for other iterations of arbitrarily selecting 5 discussions.

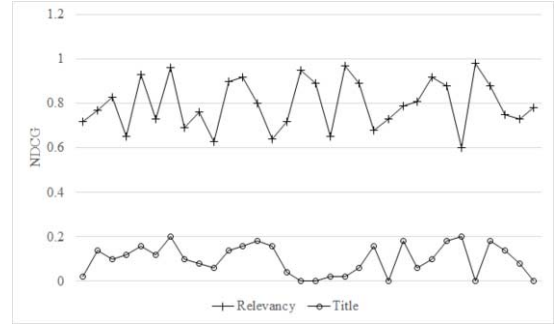


Figure 6 – NDCG for Recommended Citations

Discussion

The results show that our approach was able to retrieve citations based on forum discussion, while taking into account the relevance of keywords. Other methods for retrieving keywords in chat rooms rely on static heuristics such as a fixed time window[14]. As an example, only the last 5 messages could be used to determine the keywords for retrieving recommendations. Our method moves away from static heuristics and applies machine learning for dynamic retrieval of relevant keywords. Figure 7 shows how our neural net’s keyword selection relates to a dynamic window metric for three randomly selected Doc2Doc forum discussions over 10 incoming messages. The window is set to the number of messages from the latest to the one containing the oldest keyword selected by the neural net. Moreover, text summarization methods such as TextRank[15], and ensemble keyword extraction systems such as AlchemyAPI⁴ are not adequate for extracting keywords to summarize a forum or chat room discussion because they do not take into consideration the decay in relevance of the keywords over time.

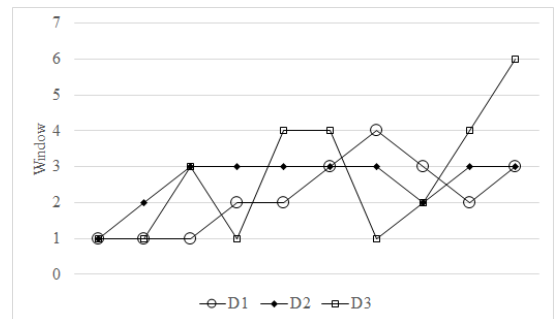


Figure 7 – Dynamic Window for Discussions D1, D2, D3

A limitation of our study is reliance on forum discussions to simulate chats, instead of actual real-time conversations. This

⁴ AlchemyAPI available at <http://www.alchemyapi.com>

impacts our trained neural net because feature properties such as *firstTime* and *lastTime* will have relatively much smaller values for real-time discussions. Another area of improvement is increasing the training dataset size, which could improve precision. Furthermore, having multiple annotations would strengthen the quality of the neural net, and also provide insights into whether dynamic time windows are based on per-user preferences.

For future work, we aim to gather data from actual real-time chat logs for medical professionals from our health portal Cardea (currently under development), where patients and medics can share experiences, ask questions, and chat in real-time within specialized areas for patient-patient, patient-medic, and medic-medic discussions. Within Cardea, the recommender system can also be used to suggest related content and citations in the discussion forums. Moreover, we intend to add feedback mechanisms to PubMedReco so users can positively or negatively vote on the recommended citations.

Conclusion

This research presented the PubMedReco recommender system which can analyze a forum or chat discussion to extract the relevant medical terms, and then query PubMed to suggest citations that are related to the ongoing discussion. PubMedReco overcomes the limitation imposed on users of online discussion environments whereby users would have to leave their chat interface to search for medical articles in a new web browser window, and formulate an appropriate search query to retrieve relevant results. We demonstrated the feasibility of PubMedReco using BMJ's Doc2Doc forum datasets. We also evaluated our system to determine the quality of its neural net's relevance predictions, the training annotations used, and the recommended citations. To the best of our knowledge, the proposed system is the first of its kind in the medical domain. Unlike other real-time chat recommender systems surveyed that use static time window heuristics, the proposed system presents a novel and dynamic machine learning approach for determining keyword relevance within health forums and chat rooms.

Acknowledgements

The authors wish to acknowledge and thank the Alberta Machine Intelligence Institute (Amii), Edmonton, Alberta, Canada, for funding and supporting this research project.

References

- [1] Z. Lu. PubMed and Beyond: A Survey of Web Tools for Searching Biomedical Literature. *J Database* (2011).
- [2] S. J. Nelson. Medical Terminologies That Work: The Example of MeSH. *IEEE International Symposium on Pervasive Systems, Algorithms, and Networks* (2009), 380-384.
- [3] Bethesda (MD): National Center for Biotechnology Information (US). Entrez Programming Utilities Help (2010). <http://www.ncbi.nlm.nih.gov/books/NBK25501>
- [4] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, editors. *Recommender Systems Handbook*. Springer (2011).
- [5] P. Jain, A. Bergen, L. Castaneda, and H. A. Muller. PALTask Chat: A Personalized Automated Context Aware Web Resources Listing Tool. *IEEE World Congress on Services* (2013), 154-157.
- [6] F. Minami, M. Kobayashi, and T. Ito. A Proposal on Recommender System Based on Observing Web-Chatting. *New Challenges in Applied Intelligence Technologies*, Springer (2008), 87-96.
- [7] W. Hong, L. Li, and T. Li. Product Recommendation with Temporal Dynamics. *Expert Systems with Applications* (2012), 12398-12406.
- [8] B. Chandramouli, J. J. Levandoski, A. Eldawy, and M. F. Mokbel. StreamRec: A Real-Time Recommender System. *ACM SIGMOD Conference* (2011), 6-8.
- [9] D. Kang, D. Han, N. Park, S. Kim, U. Kang, and S. Lee. Eventera: Real-time Event Recommendation System from Massive Heterogeneous Online Media. *IEEE International Conference on Data Mining Workshop* (2014), 1211-1214.
- [10] G. Shani and A. Gunawardana. Evaluating Recommendation Systems. *Recommender Systems Handbook* (2011), 257-298.
- [11] P. G. Campos, F. Diez, and I. Cantador. Time-Aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols. *ACM J User Modeling and User-Adapted Interaction* (2014), 67-119.
- [12] J. J. Hopfield. Artificial Neural Networks. *IEEE Circuits and Devices Magazine* (1988), 3-10.
- [13] C. D. Manning, P. Raghavan, H. Schütze. Introduction to Information Retrieval, *Cambridge University Press* (2008)
- [14] M. Montaner, B. Lopez, and J. L. De La Rosa. A Taxonomy of Recommender Agents on the Internet. *Artificial Intelligence Review* (2003), 285-330.
- [15] R. Mihalcea, P. Tarau. TextRank: Bringing Order into Texts. *Association for Computational Linguistics* (2004).
- [16] T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. *arXiv* (2013).

Address for correspondence

Hamman W. Samuel, Email: hwsamuel@ualberta.ca

Looking for the Best WOW: Understanding the Nurses' Needs

Bibiana Schachner, Zulma González, Rodrigo Cano, Daniel Luna, Sonia Benítez

Health Informatics Department, Hospital Italiano de Buenos Aires, Argentina

Abstract

The effective use of nurses' time for providing increasingly safe, efficient, and patient-centered care is a major concern for healthcare managers as well as for nurses themselves. Different solutions have been used aimed at improving those times by providing nurses with mobile and 'on wheels' alternatives for bedside care. Nevertheless, the selection of solutions is still a complicated organizational decision. This paper describes the evaluation of ergonomic characteristics of five local mobile carts for the bedside nursing care at Hospital Italiano de Buenos Aires. Cornell University's checklist was used for data collection according to five domains: car handling, work surface and data load, documents and screen reading, storage and dimensions, and energy and cleaning. Considering the scarcity of literature on needs for WOWs assessment, the findings of this paper represent a valuable approach to the requirements of nurses in real work environments and a support for decision-making based on nursing observations.

Keywords:

Computers; Nurses; Informatics

Introduction

The decrease in time that nurses use to perform tasks related to direct patient care and to document evidence of the care they provided continues to be a challenge for administrators and an unmet demand for nurses [1].

The use of computers and workstations on wheels (COWs and WOWs) is a growing trend for bedside care in healthcare settings. It is an acceptable notion that the use of time and coordination of care will be more effective if devices are used to load the data where they are generated ("point of care") [2]. However, the best solution has not yet been found and although there are recommendations to take into account while making decisions on the subject, there is little research about it [3,4].

Different health professionals can use WOWs but nurses are the main users, and they use them to record patients' assessments and to transport medications [3, 5].

Studies show that WOWs for bedside care that are available on the market have high costs and do not include mobile technology and other equipment that serve as support. They also have some other related problems, such as long inactivity times due to battery recharge, inadequate dimensions, difficult handling [6], and designs based on the devices' ergonomics—and not that of the cart or WOW itself—thus generating low adherence to use [7].

It is imperative for the development of a mobile app for nursing data load at the point of care, and for our nurses'

appraisal and needs to make an assessment and collect information regarding the characteristics of different carts and workstations on wheels. This is also useful for the selection that complies with ergonomic recommendations of these devices in nurses' bedside care environments.

Methods

Setting

The study was conducted at the Hospital Italiano de Buenos Aires (HIBA), an academic hospital founded in 1853, located in Buenos Aires, Argentina. As a JCI-accredited and stage 6+ HIMSS-EMRAM hospital, HIBA belongs to a nonprofit healthcare network including 25 outpatient centers and 150 offices located in the city of Buenos Aires. It has an infrastructure supporting 750 inpatient beds, 41 operating rooms, and a home care network with 800 beds. The organization employs nearly 2800 physicians, 1600 nurses, and 1900 employees in administrative services and management. In the past 20 years, HIBA has developed and implemented an "in house" health information system, including clinical and administrative data. The electronic health record (EHR), named ITALICA, is a modular, problem-oriented, and patient-centered system with different settings (outpatient, inpatient, emergency, and home care). It includes clinical documentation system, medication administration using bar coding (intensive care areas), and computerized provider order entry (CPOE). Sections according to the stages of the nursing care process organize the electronic nursing record. Nurses must also diagnose using NANDA-I taxonomy II diagnoses classification and care plan that is based on the nursing interventions classification (NIC).

Study design

This study was cross-sectional, observational and descriptive work, with mixed methodology.

Phase 1:

We used three workstations on wheels (named Dina, Jordan, and Andrea) from July 21 to August 15, 2016. Different local suppliers provided them according to pre-established requirements based on the various necessary characteristics. The suppliers did not provide WOWs with bar code scanners, computer or mobile devices such as a PC tablet (except for Andrea) (Figure 1), because the purpose was to evaluate the carts alone at this stage and to assess devices and barcode readers in a subsequent stage.

A convenience sample was performed. Data collection was made through observations made by the adult and pediatric general care nurses from four different wards, including five nursing shifts (morning, afternoon, night number 1, night number 2, and weekends). They were provided with a copy of

a Cornell checklist, translated into Spanish, which included ergonomic factors for computer carts in health environment. The Cornell checklist is developed by the Cornell University and represents a guide on considerations to evaluate a single cart or a WOW. Taking into account the anthropometric data of adults in the USA, the Cornell checklist is organized into five sections: a) Cart maneuvering; b) Work surfaces and data input; c) Screen reading; d) storage/accessories/power; and e) Hygiene in addition to free space for comments. The items of each section have a binary response (Yes/No) option about whether the cart satisfies the item or not. If all items are relevant, the total number of "yes" values can be added at the end of each of the five sections and the overall total score can be summarized at the end of the entire checklist. The maximum possible score is 35 and the higher score represents the computer cart with better ergonomic design. A list of activities to be carried out by the nurses with each computer cart was drawn up in the chosen areas before answering the checklist (Table 3). These activities were presented and explained to the nurse teams. Once the days established in the work plan schedule were met, the cart in a particular area was removed along with the filled checklists and a new cart was delivered to that area. Once the cart had completed its rotation in all the four sectors, it was returned to the Health Informatics Department.

The data based on the responses by the nurses for each cart by area and nursing shift were manually entered into an Excel® data sheet. We performed descriptive statistics for each cart with Stata®13. We used non-parametric Kruskal-Wallis test for independent samples to compare differences between the carts. Then we calculated a pairwise comparison of total scores sorted by each cart with the Dunn's procedure. Furthermore, comments made in the comment space by the nurses in each ward were analyzed independently by the Informatics nurses to gain additional insight about items that were possibly not in the structured checklist.

Phase 2:

Once the data were processed and analyzed, the cart with the best score was returned to the supplier for the necessary adjustments and improvements. A new provider handed over a new cart called Mariano. The WOW 'ANDREA renewed' and MARIANO were delivered to the same sectors and teams of nurses who had participated in the Phase 1 from October 20 to November 9 for following the same methodology that we used in the first phase. Then we performed descriptive statistics for each cart, and a t-student test to compare means from the two phases.

Results

A total of 59 checklists were completed during the three-week rotation of the carts during the first phase. ANDREA was evaluated 20 times, obtaining a minimum score of seven and a maximum of 34, with a mean of 15.85 points (SD 6.9). Meanwhile DINA was also evaluated 20 times, obtaining a minimum score of three and a maximum of 16, with a mean of 6.7 points (SD 3.2). On the other hand, JORDAN was evaluated 19 times obtaining a minimum score of two and a maximum of 16, with a mean of 7.3 points (SD 3.7). Figure 1 shows the carts evaluated in both the phases and Table 1 summarizes the total number of observations for each cart by sector with respective mean, the standard deviation, as well as the minimum and maximum score obtained per sector. ANDREA got the best scores in the carts observed and it was

selected for a new evaluation after its reconfiguration by the supplier. The main improvements requested from ANDREA's supplier were in the areas of the handle with a better grip, the tray and the bracket where the optical scanner would hang, and the main pillar where the internal power cables of the device and battery would fit in. The other two (DINA and JORDAN) were discarded. A Kruskal-Wallis test was performed to compare differences between carts. With a type I error of 5%, at least one of the carts had a total score different from the rest (Chi-squared (23.39 (2d.f.), $p = 0.0001$). Then, after controlling with Dunn's pairwise comparisons, we encountered statistical significant differences between ANDREA and DINA (4.66, $p < 0.00001$), ANDREA and JORDAN (4.01, $p = 0.0001$), and non significant difference between DINA and JORDAN (-0.059 , $p=0.8280$).

A total of 52 checklists were completed during the second phase. The scores obtained by 'ANDREA renewed' and MARIANO were similar in general but varied by areas. Nevertheless, the t-student test to compare the means of total scores resulted in a statistic of 0.812 (50 d.f.) and a p value of 0.4209, showing that there is no statistically significant difference for the evaluation of the characteristics evaluated by the nurses with Cornell's checklist.

Table 2 summarized the total number of observations.

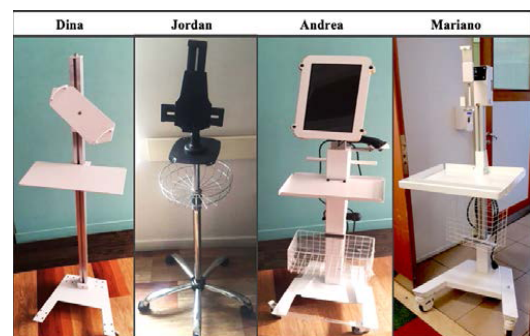


Figure 1 – evaluated WOWs

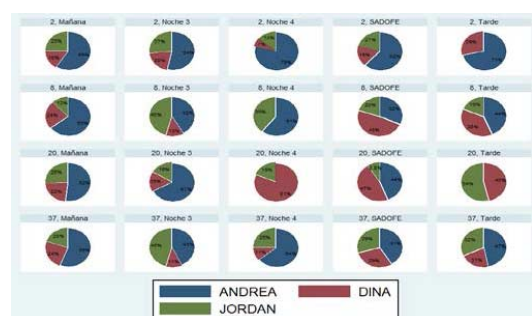


Figure 2 – Scores per cart, area and shift – phase 1

Comments analysis

We mention the themes that emerged in the first phase from nurses' comments related to the topics in the Cornell's checklist. Nursing staff in each ward identified several negative features of WOWs: lack of stability and difficulty in maneuvering (DINA), noise, unstable cart wheels, large footprint that made 'it impossible to walk beside it (JORDAN)'.

Table 1 – Assessment of Carts by Sectors and Shifts

Cart	Sector			
	2	8	20	37
ANDREA				
Observation by area	5	5	5	5
Mean	25.6	10.6	13.2	14
SD	5.12	3.91	2.04	4.18
Minimum score	21	7	11	7
Maximum score	34	17	15	18
DINA				
Observation by area	5	5	5	5
Mean	6.8	5	10	5
SD	2.58	1.41	4.41	1.58
Minimum score	3	3	6	3
Maximum score	10	6	16	7
JORDAN				
Observation by area	4	5	5	5
Mean	8.75	6.4	5.8	8.8
SD	2.62	4.33	2.68	4.71
Minimum score	6	2	3	5
Maximum score	11	11	9	16

Adult general care: sectors 2, 20 and 37. Pediatric general care: sector 8

Table 2– Assessment of 'ANDREA renewed' and MARIANO

Cart	Sector			
	2	8	20	37
'ANDREA renewed'				
Observation by area	9	1	5	9
Mean	18.67	20	16	17.78
SD	4.58		6.67	3.15
Minimum score	10	20	8	10
Maximum score	27	20	26	21
MARIANO				
Observation by area	5	4	7	12
Mean	18.6	13.5	15.28	18.16
SD	2.07	5.26	4.15	4.04
Minimum score	17	9	8	12
Maximum score	22	21	20	25

The size 'too small' or 'inadequate', adding a basket and a handles emerged as very important aspects. The main theme in all the observations was that the tray was 'inappropriate to

administer the medications', as it was very small or had no containment edges.

The theme related to the need for more comfortable handles and the basket, as well as the noise when moving the carts emerged again in the second phase. The nurses appreciated the improvements from the previous evaluation for ANDREA, but now it was 'too big to mobilize considering that patient rooms are small'. Another new element was that the power cable was 'very short'.

Table 3 – Activities List

Activities
<ul style="list-style-type: none"> ▪ Handle the cart by "pulling" the handle ▪ Adjust handle to fit comfortably ▪ Pull and push the cart, causing it to move ▪ Move the cart in a certain direction ▪ Move the cart through different areas where you walk for your work: <ul style="list-style-type: none"> ○ Nursing Station ○ Preparation of medication area (Bunker) ○ Hallways ○ Patient Room ▪ Elevate the work surface of the cart more than one meter from the floor ▪ Take notes on the work surface of the cart ▪ Place the medication tray on the work surface and move the cart to a room ▪ Adjust the inclination of the working surface of the cart ▪ Turn the work surface of the cart ▪ Lift and lower the platform for the keyboard as if to use it to write while: <ul style="list-style-type: none"> ○ Standing ○ Seating ▪ Type on the surface where the device's keyboard would go ▪ Put the items you use to do your work in the cart and move it through different areas ▪ Brake the wheels of the cart and try to move it ▪ Remove the wheel brake of the cart ▪ Throw liquid (small amount) onto the work surface and wipe it with a dressing ▪ Rotate the bracket for the device to left and right without moving the cart ▪ Rotate the bracket for the device up and down (changing the angle) without moving or turning the cart

Discussion

We performed an ergonomic assessment of five WOWs for bedside nursing care, seeking to find a solution to difficulties experienced by our nurses in collecting data at the point of care, and also for the purpose of improving communication and coordination of care. We aligned our work with suggested changes in healthcare systems to achieve more effective processes that truly support the way care is delivered and transformed [8][9].

Our evaluation examined the ergonomic characteristics available and desired in mobile stations. ANDREA showed a significant advantage over the other carts in the first phase and

renewed ANDREA continued in the race for a new evaluation. None of the carts, however, obtained the maximum score, that is, none has all the required characteristics. Some aspects continue to not meet the needs of some nurses at the start of the third phase. This could suggest a low adherence to cart use in future, even more so if we do not include the devices in this stage. Some aspects represent enhancements, such as inability to rotate the axis of the device used to document the patient care, size of the cart perceived by nurses as 'bulky' and difficult to mobilize. Such aspects limit the storage and efficient mobilization [6] in zones in which they will be deployed.

The results of this work can not be generalized since it was performed in a single hospital center. Furthermore, there was a lack of consistency in some answers—for example, user answered only a single survey in one ward in Phase 2—probably because they were self-administered by the nursing staff.

ANDREA was provided with a tablet and that feature had not been included during the previous training to the nurses. We had instructed them to ignore it, however, some observations may have been influenced by the mere presence of the device.

Two Health Informatics research interns, an Informatics nurse, and an Informatics physician agreed upon and decided on the final version of the English-Spanish translation process. The Cornell University's checklist [10] is neither externally validated nor adapted to Spanish; nevertheless, its use provided us a first approximation to the needs of nurses related to mobile workstations. We could include all the HIBA nursing shifts to strengthen our evaluation instead of using only the weekly day shifts. On the other hand, when we discarded the 'worst' carts in terms of score and reconfigured the best positioned cart, we received better response from the nurses who thought that their input was taken into account. We benefited from creating an elaborate list of activities—that simulated real-world scenarios in daily practice—before the evaluation. We are about to start a bedside care pilot study, indicating the third phase of WOW evaluation, with a new version of MARIANO renewed based on the findings and suggestions of the second phase. We also plan to incorporate a tablet PC, a bar-code scanner, and the test of the mobile app for nursing bedside care.

Acknowledgements

The authors would like to thank the nurses who participated in the evaluation study and acknowledge the collaboration between the Nursing Department and Health Informatics research.

References

- [1] Locsin R Purnell M. A contemporary nursing process. New York: Springer; 2009.
- [2] Carlson E, Catrambone C, Oder K, Nauseda S, Fogg L, Garcia B et al. Point-of-Care Technology Supports Bedside Documentation. *JONA: The Journal of Nursing Administration*. 2010;40(9):360-365.
- [3] Carlson E, Catrambone C, Oder K, Nauseda S, Fogg L, Garcia B et al. Point-of-Care Technology Supports Bedside Documentation. *JONA: The Journal of Nursing Administration*. 2010;40(9):360-365.
- [4] Andersen P, Lindgaard A, Prigmet M, Creswick N, Westbrook J. Mobile and Fixed Computer Use by Doctors and Nurses on Hospital Wards: Multi-method Study on the Relationships Between Clinician Role, Clinical Task, and Device Choice. *Journal of Medical Internet Research*. 2009;11(3):e32.
- [5] Beam E, Gibbs S, Hewlett A, Iwen P, Nuss S, Smith P. Evaluating Isolation Behaviors by Nurses Using Mobile Computer Workstations at the Bedside. *CIN: Computers, Informatics, Nursing*. 2016;34(9):387-392.
- [6] Jen M, Cho T, Rudkin S, Wong A, Almassi N, Barton E. Mobile COWs (Computer on Wheels): Hamburger or VEAL?. *Western Journal of Emergency Medicine*. 2016;17(5):527-530.
- [7] Whittemore D, Moll J. COWs WOWs, oh my! Focusing on the needs of nurses can help hospital IT departments make the best technology decisions and improve care giving at the bedside. *Health Manag Technol*. 2008 Jul;29(7):32, 34.
- [8] Institute of Medicine. *Crossing the Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
- [9] Hebda TL, Calderone TL. Informatics competencies for healthcare professionals the Technology Informatics Guiding Education Reform (TIGER) Initiative model. *Drug Metabol Drug Interact*. 2012;27(3):145-9.
- [10] Cornell University. CORNELL HEALTHCARE COMPUTER CART ERGONOMIC CHECKLIST [Internet]. 2016. Available from: <http://ergo.human.cornell.edu/Pub/AHquest/CUCompCartEval.pdf>

Lessons Learned After Redesigning a Personal Health Record

**María I. Smith, Gabriela Garcia, Mariana Simon, Lucila Bruchanski,
Julia Frangella, Janine A. Sommer, María V. Giussi, Daniel R. Luna**

Department of Health Informatics, Hospital Italiano de Buenos Aires. C.A.B.A. Argentina

Abstract

On May 2016, our institution implemented a redesign of the personal health record (PHR) with the aim of enhancing its use. The objective of this research was to know and to understand end users' opinions as regards PHR functionalities and the difficulties they have addressed while using the new PHR version. Research was based on a self administered survey, patient interviews and focus groups performed with out-patients. Topics examined: ways of access to the PHR log-in web page, frequency of use, type of device, most used functionalities, the different uses patients gave to PHR, perception as regards the redesign. This research allowed us to know the uses patients give to the PHR in this institution and to understand the difficulties they found in what refers to its re-design. This information constitutes the clue to motivate and accompany PHR users in the process of adoption of a patient portal.

Keywords: *Personal Health Records; Person-Centered Design, Consumer Health Information.*

Introduction

The Personal Health Record (PHR) has been defined as a set of computer-based tools that allow people to access and coordinate their lifelong health information and make appropriate parts of it available to those who need it [1]. They usually consist of provider-tethered applications that allow patients to electronically access health information documented and managed by a healthcare institution. Although patient portals are already being implemented, it is still unclear in which ways these technologies can influence patient care [2]. Within the aims of PHRs are communication, empowerment, portability, education, participation and self management [1;3].

To ensure the use of PHRs as a patient engaging and empowering tool regarding healthcare, patient adoption of the tool is crucial. Nevertheless, PHRs have not yet reached the levels of use expected. This fact can contribute to the delay of PHRs in reaching their potential [4]. It is reported in the literature that several issues could be obstructing PHR adoption by patients. Barriers to technology or internet access, and patients' health literacy levels are some examples [5]. Less attention has been paid to the quality of the patient-provider relationship related to portal use; this may be an important barrier to or facilitator of use [6]. Some patients had expressed their concern over whether this new technology might replace conversations between patients and physicians [7;8].

There are several studies reporting less PHR adoption among ethnic minorities and young healthy adult patients [9]. By contrast, a higher adoption is evidenced in disabled,

chronically-ill individuals and their caregivers [10;11]. At present, no clear consensus exists as the reasons of this low PHR adoption.

In May 2016, our institution implemented a redesign of the PHR with the aim of enhancing its use through different changes. Even when the hospital PHR was considered by patients as valuable and useful, their functionalities, accessibility and usability were limited. The needs of the different actors along with technological improvements and possibilities started to grow and thus, the project required a redesign process.

We applied user centre design techniques in each module, so as to improve interfaces, as one of the priorities. Other goals were to rearrange the administrative and clinical functionalities separately, and to adapt PHR to all types of devices. The redesign process was mainly focused on the flow rearrangement of each of the tasks in the working space. For this project to take place, a multidisciplinary working team was constituted. Its members were medical informatics and attending physicians, nurses, software developers and usability analysts. They agreed to work with user centre design and agile development methodology scrum approaches. By this mean, hospital patients not only would be a source of information, but also main characters of this redesign, during an iterative and incremental process in favor of continuous improvement.

The implementation was performed in a gradual manner, starting at January 2016. At first, patients could, optionally, log in to the new version and test it. Those who chose to try it also had the option to go back to the previous version. According to the registered data, between January and April 2016, few PHR users had chosen using the new version (only 10 %, from 240 000 users). Once the design, development and testing of the new version was complete, on May 2, 2016, the new PHR version was fully implemented. The choice to use the previous version of PHR was not available any more for any patient. This change result in a significant increase in the users' help-desk and support requests. From the analysis of the support needs, it came out that the difficulties in the use of the tool were mainly due to the interface changes. Taking this issue into account, we considered the necessity of carrying out inquiries with users/patients. In this context, new interviews, a survey and focus groups were performed.

The aim of this research was to know and to understand end user' opinions as regards PHR functionalities and the difficulties they have addressed while using the new PHR version. Besides, we analyzed the project lessons learned.

Methods

Settings

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit healthcare academic center founded in 1853. The HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires city and its suburban area. Between 2013 and 2014, over 45,000 inpatients were admitted to its hospitals, and there were 45,000 surgical procedures (50% ambulatory) and 3,000,000 outpatient visits. Since 1998, the HIBA has run an in-house-developed health information system, which includes clinical and administrative data. It has been recently certified by the HIMSS as level 6+ in the Electronic Medical Record Adoption Model, being the first hospital in Argentina and the second in Latin America reaching this stage. The HIBA health information Department, is in charge of the design, development, implementation and maintenance of almost all systems, including the EHR and the PHR, as well as the administrative systems [12;13].

Research carried out in 2014, described the mean age of the PHR users as 55,5 years old, being a 60,5 % female. The rate of registered users was almost 50 % while the rate of use was of 29,1 % [14].

Interviews and focus groups

This qualitative research was based on patient interviews and focus groups. Personal interviews were performed with outpatients in waiting-halls. A semi-structured guide was used for the sessions. Questions for the interview were prepared by our multidisciplinary portal implementation team (physicians, psychologists and nurses). Questions included: the different uses patients gave to PHR, frequency of use, perception as regards the redesign, etc. The recruitment for the interviews was performed in the waiting halls, with the collaboration of the different hospital administrative areas, which facilitated selection of those patients that were waiting to be attended. Each session was integrated by a facilitator and an observer.

In order to recruit participants for the focus groups, an announcement in the hospital web page and in the PHR was published. Applicants were asked to complete a form confirming attendance.

A semi-structured methodology with discussion triggered motivations was applied [15;16] Sessions were conformed by a facilitator (researcher and co-researcher) who conducted the conversation such that all participants could express their opinions concerning the proposed objectives. An observer was in charge of registering the encounters. The introduction was standardized by the researcher/co-researcher of the sessions; the aim of the encounters was explained, without naming any adjective. The focus group researcher/facilitator did not interfere with her particular cultural beliefs or preconceptions with respect to the research subject. Both the focus groups and the interviews were recorded and transcribed for further analysis. Also, an observer obtained written records at that moment in order to register non-verbal information.

This study obtained the approval of the Ethical Committee for research projects of our hospital. Once the aim of the study and the confidentiality of the given information were ensured, all participants gave their written informed consent before being included in the study. The study was conducted according to the World Medical Association Declarations of Helsinki dispositions and the clinical guidelines for best practices ICH E6.

Collection and further analysis of obtained data was carried out by two professionals. It was made by the codification and categorization of data based on a process of constant comparison according to Grounded theory [18]. Researchers reviewed the data collected, and repeated ideas, concepts or elements that became apparent were tagged with codes, which have been extracted from the data. Codes were grouped into concepts and then into categories. These categories serve as an orientation for the presentation of results in this article.

Survey

A self administered survey was conducted with a group of patients, who were selected because of the frequency of PHR use and because they had made suggestions through our support system. The survey was made by usability and qualitative testing experts, based on the need to characterize the way patients use the PHR application. Several topics were examined: ways of access to the PHR log-in web page, frequency of use, type of device, and most used functionalities. Furthermore, surveys considered use of the news section, where patients could find health information. Finally, there was a free text space left for comments. The results of the survey are expressed as a percentage.

Results

Interviews and focus groups

From June to August 2016, fifty one interviews were conducted in waiting halls of traumatology, ophthalmology, gynecology and obstetrics, internal medicine and surgery departments. Participants were from 30 to 80 years old, with an average of 65 years old.

Six focus groups with the participation of twenty six patients were also performed. The participants were from 30 to 95 years old, with an average of 68 years old.

From the analysis of the interviews and focus groups, analytical dimensions arose as follows.

The first one was related to the use patients gave to the PHR. We list the following answers in order of frequency: setting up medical appointments and its scheduling reminder functionality, asking for consultations to specialty physicians and for prescription renewals, visualization of medical bills and medication list visualization. Patients also read PHR news, social announcements, and they use the messaging system to communicate with their primary care physicians.

The second dimension was related to the different problems that came up after the redesign. We could differentiate:

- Technical issues
- Problems specifically related with the new design
- Change resistance
- A spectrum of needs that went from their need to be listened, to the preference to know those individuals who resolve their demands and complaints.

Several comments made by patients during the interviews and focus groups were textually quoted.

Technical issues

Participants named difficulties signing in at the PHR, mistakes on the PHR web page, password issues, problems with the web browser, and difficulties to understand the new interface.

- “At first, some red flags with the label: *-The user does not have access-* appeared. This means you publicited the new PHR without having the system running properly yet.”

- “I cannot enter to the PHR using the explorer browser.”

Problems related specifically to the new design

One of main reasons of discomfort with the PHR redesign was patients’ belief that several of the most used functionalities were not available any more. In fact, however, patients were merely not able to find them with the new organization.

- “I cannot find the consultations anymore.”
- “Since the format had been changed, I couldn’t send messages to my primary physician. It is hard for me to get familiar with this new format.”
- “I could not find the way to purchase medication.”
- “Before the change, I could visualize my x-rays or ultrasounds, but now, not anymore.”
- “I click, and a sign appears, *-The bill have been downloaded-* Where it appears downloaded? I do not know where it is.”

Concerning the typography size, patients found it too small. In some cases, they had difficulties understanding the icon meanings.

- “Anterior versions were of easier access. The same with the icon sizes. For example the messaging icon is smaller now.”

Another aspect that appeared was the need to enhance the visibility of certain functionalities, by placing them at the main menu with direct access. By this we mean facilitating patient access to them in a simple and agile manner.

Change resistance

taking into account that every change brings along different effects, PHR users were affected. This inquiry enabled us to understand that even if users might have regular access to information and communication technologies, a digital gap exists. It is acting as a barrier for the adoption of this tool, considering the high average age of our hospital population.

- “I do not understand computers.”
- “I want to avoid having to think. I rather prefer the PHR tell me what to do than me having to ask for the information.”
- “Simplify it. Do not try to improve those things that do not need to be improved.”
- “When the PHR was changed, I got angry because I have to adapt myself.”
- “We have to accept that is a chronological issue. I am sure that if I ask my grandson, he could log in properly.”

Other needs

An underlying finding was the need for patients to be listened to and taken into account when thinking about improving the tool, as well as the need to know those people who work behind the PHR and to be in contact with them.

- “Are you the PHR face?”
- “Now that I get to know you, next time I’ll have a problem with the PHR, I won’t get upset thinking on you.”
- “I am facing important personal problems. However, I wanted to participate in some way. I wanted to be listened.”

Survey

A Survey for 1849 patients was conducted. 668 individuals answered it completely. The response rate was of 36,12%. (Table 1)

In relation to the PHR access, 61% of the surveyed answer that they did it though the institutional web page, while 33% used the google browser.

As regards the frequency of PHR’s use, 43,71% answered they used it weekly and 42,07% monthly.

The most used functionality (Fig. 1) was test results visualization (93,41%), followed by appointment scheduling (87,43%) and messaging with primary care physicians (80,39%). Among those of median use we found: Consultation requests (61,68%) and medication purchasing (51,05%). The functionalities reported as least used were: The search of health care related information (26,35%), the family group functionality (9,28%), and others (11,23%).

The most used device was the desktop computer (69,6%), followed by notebook (40,1%), cell phone (31,44%) and tablet (15,6%).

Related to the resources utilized to obtain health care information, answers were as follow: Google (45,36%), PHR health information section (35,48%), they did not look for information on the internet (26%), institutional web page (22,46%), and others (6%).

Most users admit knowing the PHR news’ section (73,8%) and it meets their interests of their interest (75%).

After the inquiries, changes were performed. Among them we can mention the increase in the icons’ size according to the users’ requests, and the rearrangement of the application home section.

Aside from these changes to the application, different policies were established to favour communication and accompany the change. The support system was reorganized, instructional materials were sent to users and encounter spaces with users were increased.

After such changes, the support requests though the PHR and the help desks decreased. Beginning with 160 requests/day around the implementation period, this number was reduced to 30/day, most of them related to password neglect or problems with new users, from a total of 360,000 users enrolled to PHR.

Discussion

Denial, anger, bargaining, depression and finally acceptance are stages of grief during the change process. Frustration, hopelessness and anxiety are common feelings and they require acknowledgment of the underlying pain. Allowing time to adapt to the new system, improving functionality when possible and providing good training; these are important tools to help deal with the challenges of change [17].

This experience helped us to understand the vital importance of organizing and increasing specific resources during a PHR implementation as described above. These resources include: Communication, education, support and training, all addressed to patients and professionals. These were some of the lessons learned.

As regards support resources, the need to acquire a new help desk system and reorganize the requests in categories with the aim of optimizing resources is very important. The help desk tool that was available at the implementation period did not allow the tracking of the request orders resolved, help desk referral to other areas or direct contact with the user.

The option of using and testing the new version was only chosen by a few users. Taking this fact into account, it might have been worth developing a better diffusion of the change so as to attract more users during its optional use.

The literature reports significant challenges that could represent an obstacle for the adoption and effectiveness of the PHR, such as computer or internet accessibility limitations or the technical language commonly shown by PHRs[5]. This inquiry enabled us to understand that even if users might have regular access to information and communication technologies, a digital gap exists. It acts as a barrier for the adoption of this tool, considering the high average age of our hospital population.

Finally, the need for patients to be listened to is important (because of non-conformity with the attention received, delays in the medical appointments administrative issues, etc.). Users have shown interest in knowing the “PHR face”. This signifies that for patients, connecting a face and a name and personifying the application in some way, remains important in the virtual realm..

Many patients mentioned this aspect as encouraging them in the will to be part of the designers of the application.

The need to count on a platform serving as a communication source for users, allowing notification of new PHR functionalities or even contingency periods, have been evidenced by this research. Following our experience we are at present working with such a platform.

Conclusion

This research allowed us to know the uses patients gave to the PHR in this institution and to understand the difficulties they found in its re-design. This information constitutes the clue to motivate and accompany PHR users in the process of adoption of a patient portal, and its use as a tool to collaborate with their empowerment. During a process of implementation as described in this article, it becomes essential to count on different means of diffusion and training methods. Moreover, the supportive resource must be increased.

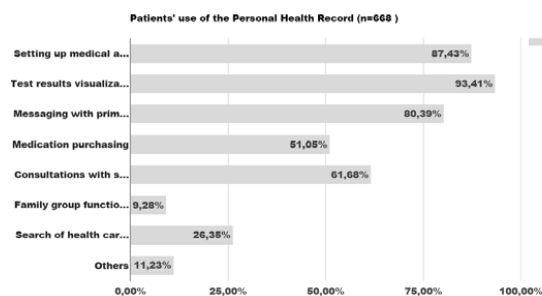


Figure 1 – Patients use of the Personal Health Record

Table 1– Survey results

Questions	%	Number of responses
How do you access to your PHR?		
Google	33,23	222
Through the institutional web page	61,08	408
Others	5,69	38
How often do you use your PHR?		
Less than 3 times a year	1,8	12
More than 3 times a year	12,43	83
Once a month	42,07	281
Once a week	43,71	292
What do you use PHR for?		
Setting-up medical appointments	87,43	584
Test results visualization	93,41	624
Messaging with primary care physicians	80,39	537
Medication purchasing	51,05	341
Consultations with specialty physicians requests	61,68	412
Family group functionality	9,28	62
Search of health care related information	26,35	176
Other	11,23	75
Which device do you usually use?		
Computer	69,61	465
Notebook	40,12	268
Cellphone	31,44	210
Tablet	15,57	104
What resources do you utilize to obtain health care information?		
Google	45,36	303
Institutional web page (HIBA)	22,46	150
PHR	35,48	237
I don't look for medical information	26,05	174
Other	5,99	40
Do you know of the PHR news section?		
Yes	73,8	493
No	26,2	175
Do you think it is interesting?		
Yes	75,15	502
No	24,85	166

Acknowledgments

We thank our patients from Hospital Italiano de Buenos Aires, who provided their time that greatly assisted the research.

References

- [1] Markle Foundation, The Personal Health Working Group Final Report, *Markle Found.* (2003) 58.
- [2] Ammenwerth, E., Schnell-Inderst, P., and Hoerbst, A., The Impact of Electronic Patient Portals on Patient Care: A Systematic Review of Controlled Trials, *J. Med. Internet Res.* **14** (2012) e162.
- [3] Kupchunas, W. R., Personal Health Record, *Orthop. Nurs.* **26** (2007) 185–191.
- [4] Greenhalgh, T., Hinder, S., Stramer, K., Bratan, T., and Russell, J.,

- Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace, *BMJ* **341** (2010) c5814–c5814.
- [5] Kutner, M., Greenberg, E., Jin, Y., and Paulsen, C., The health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy, *Education* **6** (2006) 1–59.
 - [6] Lyles, C. R., Sarkar, U., Ralston, J. D., Adler, N., Schillinger, D., Moffet, H. H., et al., Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: The Diabetes and Aging Study, *J. Am. Med. Informatics Assoc.* **20** (2013) 1128–1131.
 - [7] Hess, R., Bryce, C. L., Paone, S., Fischer, G., McTigue, K. M., Olshansky, E., et al., Original Research Exploring Challenges and Potentials of Personal Health Records in Diabetes Self-Management: Implementation and Initial Assessment, *Telemed. E-HEALTH* **13** (2007) 509–517.
 - [8] Bryce, C. L., Zickmund, S., Hess, R., McTigue, K. M., Olshansky, E., Fitzgerald, K., et al., Value Versus User Fees: Perspectives of Patients Before and After Using a Web-Based Portal for Management of Diabetes, *Telemed. E-Health* **14** (2008) 1035–1043.
 - [9] Lafky, D. B., and Horan, T. A., Prospective Personal Health Record Use Among Different User Groups: Results of a Multi-wave Study, in *Proc. 41st Annu. Hawaii Int. Conf. Syst. Sci. (HICSS 2008)*, (IEEE, 2008), pp. 233–233.
 - [10] Goel, M. S., Brown, T. L., Williams, A., Cooper, A. J., Hasnain-Wynia, R., and Baker, D. W., Patient reported barriers to enrolling in a patient portal, *J. Am. Med. Informatics Assoc.* **18** (2011) i8–i12.
 - [11] Sanders, M. R., Winters, P., Fortuna, R. J., Mendoza, M., Berliant, M., Clark, L., et al., Internet Access and Patient Portal Readiness Among Patients in a Group of Inner-City Safety-Net Practices, *J. Ambul. Care Manage.* **36** (2013) 251–259.
 - [12] González, F., Quirós, B. D. E., and Luna, D., Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires Fernán González Bernaldo de Quirós, *Cepal* (2012).
 - [13] Daniels, J., First hospital in Argentina awarded for digital maturity, (n.d.) 7–8.
 - [14] Martínez, M., Baum, A., Saldaño, A. M. G., Gomez, A., Luna, D., and De Quirós, F. G. B., Predictive variables of the use of personal health record: The Hospital Italiano de Buenos Aires study, *Stud. Health Technol. Inform.* **192** (2013) 1171.
 - [15] Scribano, A., La observación, *El Proceso Investig. Soc. Cual.* (2007) 55–69.
 - [16] R. Hernández Sampieri, C. F. C. y P. B. L., *Metodología de la investigación*, 2006).
 - [17] McAlearney, A. S., Hefner, J. L., Sieck, C. J., and Huerta, T. R., The journey through grief: Insights from a qualitative study of electronic health record implementation, *Health Serv. Res.* **50** (2015) 462–488.
 - [18] Faggiolani, C. Perceived Identity: applying Grounded Theory in Libraries. *JLIS.it.* **2** (2011).

Address for correspondence

Maria Isabel Smith, Hospital Italiano de Buenos Aires. Juan Domingo Perón 4190, C1181ACH. Ciudad de Buenos Aires. Argentina. maria.smith@hospitalitaliano.org.ar

Mobile Application for Pregnant Women: What Do Mothers Say?

Janine Sommer, Mariana Daus, María Smith, Daniel Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires

Abstract

Today, health information technologies are constantly expanding and changing, allowing more and more people to use different mobile applications to receive information and control their health condition. Based on the need to implement an application for pregnant women in the Personal Health Record (PHR) of Hospital Italiano de Buenos Aires (HIBA), an Australian survey was carried out to measure the use and utility of a pregnancy application (pregnancy app). Our results were broadly in agreement with the reference values. The survey was distributed through social networks (Facebook and Twitter) during September 2016. We obtained 235 responses from Spanish-speaking women, mostly Argentinian. In conclusion, it could be observed that a pregnancy app offers the possibility of a greater follow-up and provides reassurance to the pregnant women who use it.

Keywords:

Mobile Applications; Pregnancy; Health Records, Personal.

Introduction

Pregnancy is a period in life in which women must get used to a new state and begin to deal with new uncertainties and anxieties. The Internet allows pregnant women to access online discussion sites in order to obtain information about conception, pregnancy and maternity and facilitate communication with other women who are going through the same condition [1]. In this period patients have a great need to look for information. Prenatal education and access to information at this stage is highly valued by women, moreover when it is provided in optimal circumstances and from a safe source [2]. The popularity of pregnancy-related applications could indicate a change towards the empowerment of the patient concerning the provision of maternity care. The traditional model of “shared maternity care” in its functioning needs to incorporate electronic devices. The dependence of healthcare professionals can be reduced by the availability of interactive and personalized information delivered through an application [3].

The most important aspect an application for pregnant women must consider is meeting user needs and being reliable. There is concern regarding the poor quality of many of these applications, since they are not always based on scientific knowledge when providing information or services. Healthcare professionals need to control the quality of information provided by pregnancy applications and identify the needs of women during the prenatal period. Taking this into account we decided to implement in HIBA an application targeted to pregnant women within the Personal Health Record (PHR). The PHR of HIBA is a web-based tool where patients can take an active role in their care, have access to their own health

information and also perform certain tasks such as communicating with physicians and asking for an appointment..

In order to inquire about what pregnant women are looking for, we took and translated an Australian survey [1] that had been conducted on the use and utility of a mobile application for pregnancy with the aim of carrying it out in our context. Survey results were analyzed and taken into account for the development of an application for pregnant women within our personal health record.

Setting

Hospital Italiano de Buenos Aires is a tertiary level institution with 150 years of history. It is an academic hospital that covers the entire spectrum of health care: outpatient and inpatient care, emergencies, medical and surgical specialties, critical care, as well as home and chronic disease care. Its informatic development has been achieved from its own initiatives, which led to the appearance of multiple platforms, vocabularies and mechanisms of communication. For more than five years, the institution has been designing and building his own Hospital Information System (HIS) with the objective of linking the great diversity and variety of developments that have been developed over time within the hospital; both in the administrative and clinical layers. This resulted in the incorporation of a greater number of users who currently involve administrative staff, doctors and nurses.

During 2007 the hospital developed the Personal Health Record, a web-based tool that consists on a personal medical record that provides services and access to unified data in multiple applications. It allows patients treated in HIBA health network to interact or consult their clinical or administrative information. Self management and control of its evolution are patient’ objectives. The new PHR not only satisfies these needs, but also enables fluid and instantaneous communication with healthcare professionals. To do this, it takes advantage of the possibilities offered by the current technology and the user’s new means of communication.

Materials and Methods

So as to get to know the use and utility of mobile applications for pregnancy and parenting, an Australian survey [1] was employed. The translated survey consisted of 14 questions on mobile applications for pregnancy, the name and type of application was not asked (we excluded questions about parenting as they were not part of our goal). The survey was distributed through social networks (Facebook and Twitter) and at waiting rooms of the HIBA Women's Center during September 2016. 235 responses were obtained.

The survey had the following domains:

1. Demographic data (age, country, education level, etc.)
2. Use of mobile applications for pregnancy
3. Usefulness of mobile applications for pregnancy

Table 1 - Demographics (n=235)

Age	Country	City	Education Level
18-24: 13	Argentina: 229	CABA: 84	Middle School: 2
25-30: 63	Colombia: 1	Buenos Aires: 132	High School: 44
31-34: 69	Mexico: 1	Interior: 13	Technical: 38
35-40: 70	Costa Rica: 1	Other: 6	Incomplete degree: 55
41-45: 16	Uruguay: 1		Complete degree: 56
>45: 4	Spain: 1 France: 1		Postgraduate: 28 Associate Degree: 12

Results

83.4% of women were pregnant at the time of responding the survey. Of the 235 responses, only 152 (64.7%) women had ever used a mobile application for pregnancy and 40.1% used it frequently. Regarding the utility, 92.1% answered that they had found useful functionalities in the application they used. As a striking fact 64.5% did not check the application information sources and almost 72% did not care if the application used their given personal information.

Table 2 - Reasons for using a mobile app (respondents could select more than one option) (n = 152)

Reasons for use	%
Information on fetal development	90.1% (137)
Information about body changes	63.2% (96)
Weight gain control	15.8% (24)
Maternity forums	23.7% (36)
Reminders of shifts and medication	9.2% (14)
Keeping a pregnancy journal	23% (35)
Upload and save studies	1.3% (2)
Upload and save photos	12.5% (19)
Other	9.2% (14)

Table 2 shows the reasons that motivate respondents to use a mobile pregnancy app. The most popular reason for using a pregnancy app was to receive information on fetal development (90.1%). Information on changes in a woman's body during gestation (63.2%) was another frequent reason for its use. This is consistent with the results reported in Australia (86% and 71% respectively). The remaining reasons included control of weight gain (15.8%), discussion forums with other pregnant women (23.7%), reminders of medical shifts (9.2%), keeping a pregnancy diary (23%) and upload of pictures or ultrasound studies (1.3%). In the category "other" (9.2%) we found the functionality of the contraction counter as a reason for its use. As shown in Table 3, respondents found useful information (84.9%) that allowed them to monitor fetal development (57.9%) or changes in their body (39.5%) as well as providing

reassurance (28.3%). Also they used it to get in touch with other pregnant women (13.8%) and to store photos and videos (13.2%). Among the less useful features were the appointment reminders (7.2%), the sharing of information with family and friends (6.6%) and the storage of ultrasounds (0%).

Table 3 - Usefulness found in a mobile pregnancy app (respondents could select more than one option) (n = 152)

Usefulness	%
Provides Information	84.9% (129)
Allows monitoring of fetal development	57.9% (88)
Allows monitoring of body changes	39.5% (60)
Provides reassurance	28.3% (43)
Allows communication with other pregnant women	13.8% (23)
Avoids forgetting shifts and medical details	7.2% (11)
Allows storage of photos and videos	13.2% (20)
Allows storage of ultrasounds	0% (0)
No useful functionalities are found	3% (2)
Other	3.9% (6)

Discussion

Several studies have shown that pregnant women are using the Internet to look for information related to pregnancy, and that online forums are spaces they use for support and guidance. Researchers are exploring how new technologies, including internet forums, websites, email, YouTube, text messaging and smartphones, can be used to promote maternal and newborn health. Research suggest that these technologies are a new and promising mean for health education and communication [5]. Mobile applications allow patients to record the health information collected at each appointment and track their health. As patients interact with this information, they may become more involved in their care. Thus, patients believe that their role as patients is important, that they have the confidence and knowledge to take action, and that they can adopt behaviors to maintain and improve their health; patients can be able to take actions, ask questions to their healthcare provider, and participate in the decision making process about their treatment, being therefore partners with their provider in their health care. Patient engagement is a factor that affects the interpersonal processes of prenatal care. Interpersonal care processes conceptually and operationally cover three dimensions of clinical care: communication, patient focus on decision making, and interpersonal communication style [6]. With this in mind, we focus on this population to meet their information needs. For the development of the mobile application for pregnant women within the Personal Health Record we first investigated other experiences on the use of mobile applications. Once we found the survey conducted in Australia, and realize that it followed our same research goal, we decided to adapt it locally and apply it in our context in order to draw conclusions and using it as a guidance for the development of the application's functionalities. We do not find similar experiences regarding this issue in South America, this suggests that although it is demonstrated that mobile applications bring information to the pregnant women and empower them, there is currently lack of evidence and research on this topic in our population.

Regarding the conduction of the survey, social networks employment broadly facilitated survey diffusion. The respondents participation was voluntary and demographic characteristics were heterogenous, which gave us varied and rich opinions. In parallel we conducted the survey in person at the waiting area of the Women's Center of the Hospital Italiano de Buenos Aires. A total of 235 responses were received; 15% of surveys were conducted in person. Results showed a wide range of app use, from those who have never utilized an application to those who use it every day. Other relevant information such as the importance they gave to the source of information they were consulting, or if they were concerned about the sensitive personal information storage and management within the application was also obtained.

Developing an application according to the needs of the end user is extremely important [7]. Pregnancy is a unique life experience that evokes a range of emotions from great joy and anticipation to paralyzing anxiety. This increase on emotions facilitates a greater demand of experience exchange with others as well as instant connectivity, professional consultation and reassurance needs. The panacea to enjoy and relief from these emotions could be the smartphone and its associated applications. These devices could potentially allow women to communicate, follow the progress of their pregnancy, and relieve distress interactively [3]. Informational applications, which are non-interactive and are in the base of reference, constitute the largest category of applications related to pregnancy. These applications cover a range of topics related to maternal and fetal health, from general pregnancy to more specific information, such as the mother's diet for gestational diabetes. In many ways, their quantity and popularity suggest that there is a significant demand for such information [4]. From this perspective, this study was very significant as it identified the patterns of use and the need of applications in pregnant women, as well as clarifying other aspects that should also be taken into account. This resulting considerations were taken as the basis for the development of a specific application adapted to the needs of women in this condition.

The objective is that pregnant woman can find a tool tailor-designed for her that provides with pertinent information, to carry out a collaborative accompaniment of the healthcare team from the moment of conception through pregnancy period and till childbirth; and provide useful tools that allow users to record moments that they consider as important in a "pregnancy journal." This application will be linked to our PHR created as a tool to empower patients and give them the capability to participate in their own health care.

The access to the application within the same PHR gives the patient the reliability that the offered health contents are created and validated by specialized professionals and endorsed by our institution. The application will be called *Personal Portal for Pregnant Women*. Its will contain three time lines differentiated by color, with different types of content: clinical (appointments with professionals, medication to be taken), contextual information (development of the baby according to gestational age, changes in the mother and studies to be done in each trimester) and information added by the patient (notes, reminders, photos). The pregnant woman can monitor her pregnancy through the platform, in real time. (Figures 1, 2 and 3). This project, not implemented yet, it is in the last phase of development. Once fully implemented, we will measure its impact as a future line of research.

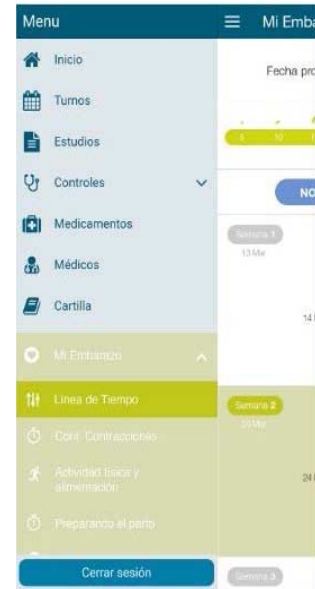


Figure 1 - Screenshot of the mobile version prototype of the homepage menu

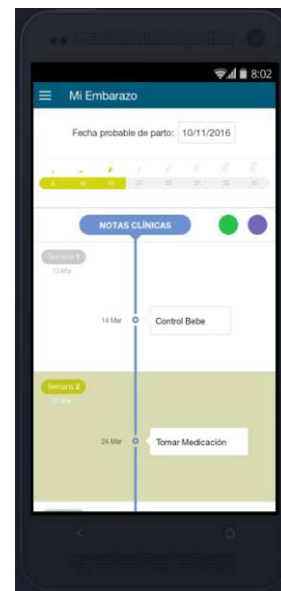


Figure 2 - Screenshot of the mobile version prototype



Figure 3 - Screenshot of the mobile version prototype of the blue timeline (clinical data)

Conclusions

The survey used allowed us to get to know the reasons why women use a mobile application during pregnancy. Although we already had information related to the Australian population, we wanted to characterize the population in our geographical area. In conclusion, we could observe that the pregnant women motivations for the use of a mobile application during pregnancy are highly concordant with those described in

the reference paper [1]. This allowed us to define the functionalities to develop a mobile application for pregnant women within the HIBA PHR.

References

- [1] D. Lupton and S. Pedersen, An Australian survey of women's use of pregnancy and parenting apps, *Women and Birth* **29** (2016), 368-375.
- [2] M.L. Nolan, Information giving and education in pregnancy: a review of qualitative studies, *The Journal of perinatal education* **18** (2009), 21-30.
- [3] N. Tripp, K. Hainey, A. Liu, A. Poulton, M. Peek, J. Kim, and R. Nanan, An emerging model of maternity care: Smartphone, midwife, doctor?, *Women and Birth* **27** (2014), 64-67.
- [4] Y. Lee and M. Moon, Utilization and content evaluation of mobile applications for pregnancy, birth, and child care, *Healthcare informatics research* **22** (2016), 73-80.
- [5] D. Rodger, A. Skuse, M. Wilmore, S. Humphreys, J. Dalton, M. Flabouris, and V. Clifton, Pregnant women's use of information and communications technologies to access pregnancy-related health information in South Australia, *Australian Journal of Primary Health* **19** (2013), 308-312.
- [6] C.J. Ledford, M.R. Canzona, L.A. Cafferty, and J.A. Hodge, Mobile application as a prenatal education and engagement tool: a randomized controlled pilot, *Patient education and counseling* **99** (2016), 578-582.
- [7] K. Chyjek, S. Farag, and K.T. Chen, Rating pregnancy wheel applications using the APPLICATIONS scoring system, *Obstetrics & Gynecology* **125** (2015), 1478-1483.

Address for correspondence

janine.sommer@hospitalitaliano.org.ar

Operational Data Model Conversion to ResearchKit

Iñaki Soto-Rey, Sophia Geßner, Martin Dugas

Institute of Medical Informatics, University of Münster, Germany

Abstract

The increasing use of electronic health information systems brings up an unresolved issue: the lack of interoperability between these systems. The Clinical Data Interchange Standards Consortium's Operational Data Model (ODM) is an xml based standard for the exchange of clinical data and metadata. The University of Münster has been using ODM to store medical forms in a web based metadata registry called Portal of Medical Data Models, which includes a complete set of tools to transform ODM forms into other formats. One kind of medical form is the Patient Reported Outcome, a trending type due to its easy integration with mobile data capture systems. ResearchKit is a development framework that allows the easy creation of these for iOS devices; unfortunately its current interoperability is limited. This research proposes a mapping between ODM and ResearchKit and presents the successful implementation of a converter for ODM into JSON based ResearchKit readable files.

Keywords:

Patient Reported Outcome Measures, Metadata, Health Information Systems.

Introduction

With the increase of electronic health information systems (HIS) in place, a problem arises: the interchange of information between these, formally called "interoperability". This is still one of the biggest issues in Medical Informatics despite of the several initiatives pursuing a solution [1]. One of the most recognised institutions working on this field is the Clinical Data Interchange Standards Consortium (CDISC), a non-profit organisation which aims to enable HIS interoperability through vendor-neutral, platform-independent and freely available data standards [2].

The CDISC Operational Data Model (ODM) proposes a format for exchanging and storing clinical data and metadata. This xml based standard is compliant with the American Federal Drug Administration (FDA) submission guidelines for metadata [3]. ODM is widely spread and it is supported by numerous HIS. Hume et. al. reviewed 69 publications that describe several uses of ODM [4].

An example of the use and potential of ODM is the Medical Data Models (MDM) web portal¹. The MDM project started in 2011 at the University of Münster with the purpose of enabling form based metadata exchange [5]. The web based portal developed by the Münster team contains, to date, more than 250,000 items encapsulated in more than 10,000 ODM medical forms, together with the so-called "ODMToolbox", a set of tools to generate, modify and transform ODM forms. The portal allows users to easily visualise, rate, comment and down-

load medical forms, originally created, semantically annotated (using Unified Medical Language) and stored as ODM XML files, in several languages and formats such as PDF (with and without comments), CDA, CSV, FHIR (as JSON and XML format), Macro-XML, REDCap, SQL, SPSS, ADL, R and XLSX [6]. Furthermore, the portal also includes tools such as the so-called "ODMedit", a tool to create medical forms in ODM; once the forms are created, they can either be downloaded or directly uploaded to the MDM portal [7] and the "ODMSummary", which allows the comparison of multiple semantically annotated medical forms in a web based tool with a high usability score [8]. In constant evolution, new forms and functionalities are regularly being included to the MDM portal, which became in 2015 the world's largest open-access metadata registry.

One popular type of medical form are Patient Reported Outcomes (PROs), defined by the FDA as "measurements of any aspect of a patient's health status that come directly from the patient (i.e., without the interpretation of the patient's responses by a physician or anyone else)" [9]. PROs are useful for the better understanding of clinical cases and the detection of patients' life quality issues, but their use and interpretation must be carefully handled [10]. Regarding their benefits for clinical research, PROs have proven to be helpful when assessing patient trial eligibility and the efficacy of new treatments and therapeutic procedures [11]. Traditional paper based collection of PROs is though cumbersome and prone to error when, for example, the collected data needs to be manually introduced in a HIS.

Electronic HIS and especially mobile based HIS offer a great opportunity for the collection of PROs. Electronic PRO systems (ePROs) lead to more accurate and complete data, improved protocol compliance, avoidance of secondary data entry errors, easier implementation of skip patterns, less administrative burden, high user acceptance, reduced sample size requirements, and potential cost saving when compared to paper based collection of PROs [12].

As several ePROs are being developed around the world, the second most important mobile device operating system after android, iOS [13], has started its own initiative to support the creation of new ePROs with a development framework: ResearchKit² (RK). RK allows developers to easily create ePROs with a standard user interface, optimised for all iOS based mobile devices. RK can be freely downloaded from GitHub³ (an online repository for open source projects) [14] and built with Xcode⁴. RK supports not only the collection of classic PROs containing traditional PRO questions (multiple choice, free text, numerical, etc.) but also other functions like a utility to provide information about the study and give consent re-

¹ <http://medical-data-models.org/>

² <http://researchkit.org/>

³ <https://github.com/>

⁴ <https://developer.apple.com/xcode/>

motely; or information collection from several tasks available in mobile devices such as location, reaction time, balance, distance walked and multi-media records (audio, photo and video). These innovative measurements open a new spectrum of possibilities for clinical data collection and research, some of which are already starting to be used in clinical studies [15].

Unfortunately, the current version of RK does not support clinical standard formats, which hampers the exchange of information with other systems and the automatic inclusion of PROs, by for example re-using medical forms from repositories such as MDM.

The purpose of this research is to investigate the feasibility of a transformation from ODM forms into RK questionnaires and the development of a converter between these two formats, which should be integrated within the ODMToolbox of the MDM portal.

Methods

The first step was the investigation of the RK application programming interface (API) and the suggestion of a mapping between ODM and RK. For this, the ODM 1.3.2 specification was taken as reference and a list of the elements conforming ODM's hierarchy was built. With this list, matches for every element were defined based on RK's API and the RK items best representing ODM's attributes and classes were listed next to their ODM homologs.

The second step involved the development of the ODMtoResearchKit converter, which was fulfilled using Java programming language and NetBeans⁵ as integrated development environment. During the development, some of the initially suggested mapped items were corrected.

The versions of the standard and software supported by the converter are ODM v1.3.2 and RK on its last update on GitHub (to date 853 forks), together with AppCore (81 forks), a model built on top of RK that includes important features not initially included in RK such as: dashboard with progress graphs; data storage back end; JSON serialization and deserialization and integration with Sage Bionetworks' Bridge service [16]. For the manipulation of ODM forms, a Java library developed by the University of Münster in 2014 was used [17]. This library allows users to generate an ODM Java object from an XML based ODM form, which facilitates its manipulation with a Java based program for its conversion into other formats.

The functionality tests were performed using ODM samples from the MDM portal, converting them into RK JSON and importing them with the sample application available in RK's source code and the open source application developed by the Icahn School of Medicine at Mount Sinai: "Asthma Health"⁶, which uses RK and AppCore and is able to import JSON files as questionnaires.

Results

The result of the mapping between ODM forms and a RK based questionnaires can be visualised in Table 1.

Table 1– Mapping between ODM elements and their equivalent ResearchKit questionnaire items

ODM Object	ResearchKit Mapping
Study	ORKInstructionStep
OID	ORKInstructionStep.Identifier
StudyName	ORKInstructionStep.Title
StudyDescription	ORKInstructionStep.Text
StudyEventDef	<i>Not applicable</i>
Name	<i>Not applicable</i>
OID	<i>Not applicable</i>
Repeating	<i>Not applicable</i>
Description	ORKFormStep.Title
ItemRef mit	ORKFormStep.formItems
ItemOID	
Mandatory	ORKFormStep.optional
ItemDef	ORKFormItem
DataType	ORKFormItem.answerFormat
text	ORKTextAnswerFormat
Length	ORKTextAnswerFormat.maximumlength
integer	ORKQuestionTypeInteger (ORKNumericAnswerFormat)
float	ORKQuestionTypeDecimal (ORKNumericAnswerFormat)
GE	ORKNumericAnswerFormat.minimum
LE	ORKNumericAnswerFormat.maximum
MeasurementUnit	ORKNumericAnswerFormat.unit
date	ORKDateAnswerFormat
YYYY-MM-DD	ORKDateAnswerFormat.style
GE	ORKDateAnswerFormat.minimumDate
LE	ORKDateAnswerFormat.maximumDate
time	ORKTimeOfDayAnswerFormat
datetime	ORKQuestionTypeDateAndTime
intervalDatetime	ORKTimeIntervalAnswerFormat
boolean	ORKBooleanAnswerFormat
Integer with	ORKTextChoiceAnswerFormat
CodeListItems	
ORKQuestionTypeSingleChoice	ORKTextChoiceAnswerFormat.style
CodeListRef mit	ORKTextChoiceAnswerFormat.textchoices
CodeListOID	ORKFormItem.identifier
OID	
Description	<i>Not applicable</i>
Question	ORKFormItem.text
Alias	<i>Not applicable</i>
mandatory	ORKFormItem.optional task (id<ORKTask>)
CodeListItem	ORKTextChoice
CodedValue	ORKTextChoice.value
TranslatedText	ORKTextChoice.text oder ORKTextChoice.detailText
ItemDef	ORKStep
OID	ORKStep.identifier
	ORKStep.restorable
Mandatory	ORKStep.optional
Name	ORKStep.title
Question	ORKStep.text
FormDef	ORKStep.task

The RK question types not visualised in the table and the ODM elements marked with "*Not applicable*" could not be mapped and will be analysed in the discussion.

⁵ <https://netbeans.org/>

⁶ <http://apps.icaohn.mssm.edu/asthma/>

The converter developed contains two main functionalities: One enables users to upload an ODM file in a very basic webpage deployed locally using Tomcat⁷. This basic system imports the (ODM) file and generates a JSON file that is stored locally and visualised on the webpage. This functionality was mainly developed and used for local tests. The second and main functionality is the integration within the previously introduced ODMToolbox of the MDM portal. Figure 1 represents the common workflow to use the ODMtoResearchKit converter: (1) A registered user of the MDM portal searches for a desired medical form using the MeSH-based MDM searching tool and accesses to the form. (2) The user hovers over “Download” on the left part of the display and selects “ResearchKit (JSON)”. The downloaded .zip file contains the medical form in RK JSON format, as well as the license information and a JSON document with the form’s metadata. (3) The user copies the downloaded form into the resources folder of the RK-based ePRO and includes the file in the source code.

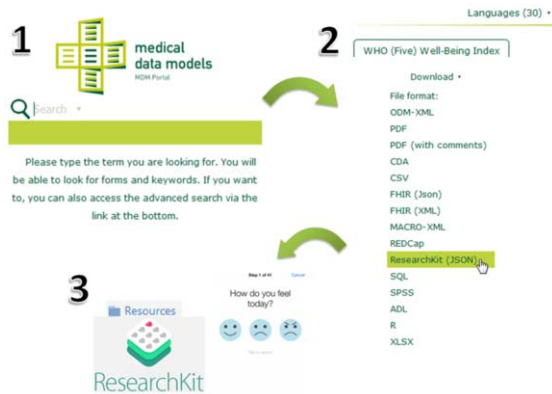


Figure 1 – Process steps to be followed in order to find a desired data model in MDM, download it in RK JSON format and include it in a RK-based ePRO.

The ODM to RK converter generates JSON files with a structure readable by any application that includes both RK and AppCore (Figure 2). In RK JSON, a questionnaire is defined as a task (RKTask). A task includes one to several questions named “steps”, as well as an identifier and a name. The steps are defined as single JSON objects and are encapsulated in a JSON array. Every step contains an identifier, a question text and an answer format and, depending on the answer format, other elements such as range limits for numerical questions or answer choices for multiple answer choice questions. It is also possible to decide whether a question has to be answered or not with the option “optional task”, which can also be imported from ODM’s “mandatory” field.

```
{
  "identifier": "F.0000",
  "name": "WHO (Five) Well-Being Index",
  "_class": "RKTask",
  "steps": [
    {
      "identifier": "I.1001",
      "question": "I have felt cheerful and in good spirits",
      "answerFormat": {
        "options": [
          {
            "shortText": "All of the time",
            "_class": "RKAnswerOption"
          },
          {
            "shortText": "Most of the time",
            "_class": "RKAnswerOption"
          },
          {
            "shortText": "More than half of the time",
            "_class": "RKAnswerOption"
          },
          ...
        ]
      }
    }
  ]
}
```

Figure 2 – Example of the first question and three first answers of the Well-Being Index questionnaire in ResearchKit JSON format, generated using the developed converter.

As RK does not support multilingual questions yet, the language of the items converted from MDM portal into RK JSON files will be the language selected for the visualisation of the form in the MDM portal. Multilingual functions for RK have been suggested and are part of the development roadmap.

The converter was positively tested using several PROs contained in the MDM portal, downloading them as RK JSON files and importing them with Asthma Health, a RK-based application available online, that includes AppCore to import RK JSON files [18].

Discussion

The converter developed is able to generate RK readable JSON files from an ODM file. With this tool, the more than 130 PROs and 9.000 medical forms available in the MDM portal can be easily downloaded and imported by RK-based apps. Besides, the test suite server developed is useful for local functionality testing and could be re-used by future converters.

Some limitations need to be pointed out: RK does not include or use a standard for using medical forms; it just provides a framework that facilitates the development of new ePROs. Thus, a mapping between ODM and the RK questionnaire structure is not straightforward. Likewise, ODM is not PRO oriented and it does not include some useful question types supported by RK.

These are the ODM item types not supported by RK:

- hexBinary | base64Binary | hexFloat | base64Float
- partialDate | partialTime | partialDatetime
- durationDatetime
- incompleteDatetime | incompleteDate | incompleteTime
- URI.

RK question types not supported by ODM:

- ContinuousScale

⁷ <http://tomcat.apache.org/>

- Email
- HealthKitCharacteristicType
- HealthKitQuantityType
- Location
- TextScale
- Multi-media (Audio, video and photo)

Together with the activities supported by RK (and not by ODM):

- CountdownStep
- Fitness
- TowerOfHanoiStep
- WalkingTask

ODM v1.3.2 includes a possibility to represent multiple-choice questions using code lists. On the other hand, it cannot be defined if several of the choices contained in a code list could be answered. This represents a limitation when converters from ODM metadata into PRO based formats are developed, as it is an essential characteristic of classic PRO questionnaires. In practical terms, this means that a conversion between ODM and PRO formats requires post-conversion handling for this kind of questions.

RK is meant to be used by ePROs created for a single study. Thus, an element that represents a study within the JSON structure is not needed (as there will only be one). ODM on the other hand, includes a great amount of information about the study that, due to RK's limitation, will be lost when generating RK compliant JSON files with this converter.

RK includes mainly user interface capabilities; it needs AppCore for useful features such as serialisation and deserialisation of JSON files. Although AppCore functionality is soon to be included in RK, to date, a RK-based ePRO would need AppCore in order to import JSON based questionnaires.

There are several systems available for the electronic collection of PROs. The most simple ones are electronic online survey systems such as LimeSurvey [19], which offer a wide variety of languages and configuration options for the easy creation and customisation of surveys, but are normally just used for anonymous collection of simple data as data privacy may be an issue. Other systems are also multi-language and web-based, but are so far only developed to be used in a clinical setting, restricting their use with the a web clip installed in a mobile device [20]. Some other ePROs just provide an application developed compliant with several operating systems and devices, but they are normally developed for a single purpose [21], [22]. Another notorious example is REDCap: a methodology and a software solution for the rapid development and deployment of electronic data capture tools to support clinical and translational research though intuitive electronic case report forms and metadata import functions [23]. Other projects like C3PRO propose a working framework for the exchange of information between ePROs and different HIS though medical standards [24], but they are focused on clinical data, and not on metadata and import of medical forms. The great availability of systems in place should be taken into account when a new one is requested.

RK is only compliant with iOS devices. Another project that uses RK as basis and provides a similar framework for An-

droid devices is ResearchStack⁸. We decided to develop firstly the ODM to RK converter, since RK represents the original framework compared to ResearchStack. Furthermore there are currently more initiatives using ResearchKit to conduct research and, most importantly, after a short look into ResearchStack's source code, one can observe that ResearchStack has preserved RK's structure and functionalities, which suggests that the JSON files generated by the ODMtoResearchKit converter could also be importable by ResearchStack based ePROs.

Electronic PROs have surpassed the possibilities of PROs. Modifications on traditional PROs should be considered so that they include new question types and tasks available only on electronic devices (previously mentioned). To the knowledge of the authors, there are no official guidelines for these question types and tasks to be included as part of PROs yet. Likewise, ODM should include some of these useful question types and tasks in order to update it to the new ePROs, as well as a solution for the multiple choice questions previously explained. RK should consider the possibility to select a different graphic interface and survey languages. As soon as this functionality is available, a small update on the ODMtoResearchKit converter needs to be carried out so that the JSON files generated contain the elements in various languages. The MDM portal contains, to date, more than 130 PROs. This number is thought to be low compared to the total amount of data models included in the portal. The most important reason for this is the lack of licence free PROs. A feature to include non-licence free PROs in the MDM portal could be developed so that only users with rights to use these kind of PROs could download them. The converter could be enhanced with a library to manipulate RK objects in Java, which would facilitate the export in different formats (swift or objective-c code for example) and further tests are needed to demonstrate the compatibility with ResearchStack.

Conclusion

A mapping between ODM and RK questionnaires was suggested and a converter for ODM forms into RK readable JSON forms developed. With this converter integrated within the MDM portal, more than 130 PROs and 10,000 medical forms contained in it can be easily imported by RK-based ePROs. Further work should include a Java based RK questionnaire library and tests to demonstrate the generated JSON files' compliance with ResearchStack.

Acknowledgements

This work was supported by German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant DU 352/11-1).

References

- [1] B. Dixon, *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*. Academic Press, 2016.
- [2] 'CDISC Mission and Principles', CDISC, 27-May-2009. [Online]. Available: <https://www.cdisc.org/about/mission>. [Accessed: 23-Nov-2016].
- [3] V. Huser, C. Sastry, M. Breymaier, A. Idriss, and J. J. Cimino, 'Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM)', *J. Biomed. Inform.*, Jul. 2015.

⁸ <http://researchstack.org/>

- [4] S. Hume, J. Aerts, S. Sarnikar, and V. Huser, 'Current applications and future directions for the CDISC Operational Data Model standard: A methodological review', *J. Biomed. Inform.*, vol. 60, pp. 352–362, Apr. 2016.
- [5] B. Breil et al., 'Multilingual Medical Data Models in ODM Format: A Novel Form-based Approach to Semantic Interoperability between Routine Healthcare and Clinical Research', *Appl. Clin. Inform.*, vol. 3, no. 3, pp. 276–289, 2012.
- [6] M. Dugas et al., 'Portal of medical data models: information infrastructure for medical research and healthcare', *Database J. Biol. Databases Curation*, vol. 2016, Feb. 2016.
- [7] M. Dugas, A. Meidt, P. Neuhaus, M. Storck, and J. Varghese, 'ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository', *BMC Med. Res. Methodol.*, vol. 16, Jun. 2016.
- [8] M. Storck, R. Krumm, and M. Dugas, 'ODMSummary: A Tool for Automatic Structured Comparison of Multiple Medical Forms Based on Semantic Annotation with the Unified Medical Language System', *PloS One*, vol. 11, no. 10, p. e0164569, 2016.
- [9] U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research, U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research, and U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health, 'Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance', *Health Qual. Life Outcomes*, vol. 4, p. 79, Oct. 2006.
- [10] J. Greenhalgh, 'The applications of PROs in clinical practice: what are they, do they work, and why?', *Qual. Life Res.*, vol. 18, no. 1, pp. 115–123, Feb. 2009.
- [11] P. R. Deshpande, S. Rajan, B. L. Sudeepthi, and C. P. Abdul Nazir, 'Patient-reported outcomes: A new era in clinical research', *Perspect. Clin. Res.*, vol. 2, no. 4, pp. 137–144, 2011.
- [12] S. J. Coons, S. Eremenco, J. J. Lundy, P. O'Donohoe, H. O'Gorman, and W. Malizia, 'Capturing Patient-Reported Outcome (PRO) Data Electronically: The Past, Present, and Promise of ePRO Measurement in Clinical Trials', *The Patient*, vol. 8, no. 4, pp. 301–309, 2015.
- [13] K. P. C. B. www.kpcb.com, '2016 Internet Trends Report'. [Online]. Available: <http://www.kpcb.com/internet-trends>. [Accessed: 19-Dec-2016].
- [14] 'ResearchKit/ResearchKit', GitHub. [Online]. Available: <https://github.com/ResearchKit/ResearchKit>. [Accessed: 14-Dec-2016].
- [15] E. R. Dorsey, Y.-F. Yvonne Chan, M. V. McConnell, S. Y. Shaw, A. D. Trister, and S. H. Friend, 'The Use of Smartphones for Health Research', *Acad. Med. J. Assoc. Am. Med. Coll.*, Apr. 2016.
- [16] 'ResearchKit/AppCore', GitHub. [Online]. Available: <https://github.com/ResearchKit/AppCore>. [Accessed: 15-Dec-2016].
- [17] I. Soto-Rey, M. Dugas, and M. Storck, 'Implementation of an ODM and HL7 Compliant Electronic Patient-Reported Outcome System', *Stud. Health Technol. Inform.*, vol. 228, pp. 421–425, 2016.
- [18] 'Asthma Health App | Apple ResearchKit | Icahn School of Medicine at Mount Sinai', *Asthma Health App | Apple ResearchKit*. [Online]. Available: <http://apps.icaahn.mssm.edu/asthma/>. [Accessed: 14-Dec-2016].
- [19] 'LimeSurvey - the most popular FOSS survey tool on the web'. [Online]. Available: <https://www.limesurvey.org/>. [Accessed: 13-Mar-2016].
- [20] M. Storck, B. Trinczek, M. Dugas, and F. Fritz, 'Towards a trial-ready mobile patient questionnaire system', *Stud. Health Technol. Inform.*, vol. 205, pp. 768–772, 2014.
- [21] A. V. Bennett, R. E. Jensen, and E. Basch, 'Electronic patient-reported outcome systems in oncology clinical practice', *CA. Cancer J. Clin.*, vol. 62, no. 5, pp. 336–347, Sep. 2012.
- [22] K. Schick-Makaroff and A. Molzahn, 'Strategies to use tablet computers for collection of electronic patient-reported outcomes', *Health Qual. Life Outcomes*, vol. 13, Jan. 2015.
- [23] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, 'Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support', *J. Biomed. Inform.*, vol. 42, no. 2, pp. 377–381, Apr. 2009.
- [24] P. B. Pfiffner, I. Pinyol, M. D. Natter, and K. D. Mandl, 'C3-PRO: Connecting ResearchKit to the Health System Using i2b2 and FHIR', *PloS One*, vol. 11, no. 3, p. e0152722, 2016.

Address for correspondence

Inaki.SotoRey@uni-muenster.de

A Secure Architecture to Provide a Medical Emergency Dataset for Patients in Germany and Abroad

Michael Storck^a, Jan Wohlmann^b, Sarah Krudwig^c, Alexander Vogel^d,
Judith Born^b, Thomas Weber^e, Martin Dugas^a, Christian Juhra^b

^a Institute of Medical Informatics, University of Muenster, Muenster, Germany,

^b Office for eHealth, University Hospital Muenster, Muenster, Germany

^c CompuGroup Medical Deutschland AG, Koblenz, Germany

^d DATATREE AG, Duesseldorf, Germany

^e Institute of Aerospace Medicine, German Aerospace Center, Cologne, Germany

Abstract

The ongoing fragmentation of medical care and mobility of patients severely restrains exchange of lifesaving information about patient's medical history in case of emergencies. Therefore, the objective of this work is to offer a secure technical solution to supply medical professionals with emergency-relevant information concerning the current patient via mobile accessibility. To achieve this goal, the official national emergency data set was extended by additional features to form a patient summary for emergencies, a software architecture was developed and data security and data protection issues were taken into account. The patient has sovereignty over his/her data and can therefore decide who has access to or can change his/her stored data, but the treating physician composes the validated dataset. Building upon the introduced concept, future activities are the development of user-interfaces for the software components of the different user groups as well as functioning prototypes for upcoming field tests.

Keywords:

Emergency treatment; Emergency Service, Hospital; Health information exchange

Introduction

Due to increasing specialization and fragmentation of medical care, as well as an increased mobility of patients, the information flow, which enables a fast exchange of lifesaving information about a patient's prior medical history, is significantly hampered [1,2]. Especially in cases of emergency, where the patient is possibly unconscious and cannot be asked about previous illnesses or current medication, this information can be exceedingly vital [2,3].

To establish an environment of quickly accessible patient information in case of emergency, many countries have started to implement or plan to implement Medical Emergency Datasets (MED). Scheduled for 2018, the German Electronic Health Card (EHC) is supposed to hold a MED [4,5,6]. Authorized health professionals (e.g. primary care physicians) are asked to create these datasets, if requested by the patient. Vital information about previous diagnoses, medications, allergies and other emergency-relevant information are stored on the card [5,7]. Additionally, the patient has the opportunity to save information about the locations of important personal statements such as organ and tissue donor ID, end of life decision or a power of attorney regarding medical care.

Against the background of this upcoming implementation, the German Medical Association [8] and the University Hospital Muenster have conducted a validation study in 2014, which surveyed the usability, the necessary instructions as well as analyses on the expected usefulness of paper-based MEDs. Results indicate that the questioned medical personnel, who included emergency physicians, clinicians and paramedics, see MEDs as very useful or useful. Specifically, preliminary diagnoses and medications were rated as highly relevant for the emergency care [9]. Building upon the aforementioned findings, the pilot study 'NFDM-Sprint' is currently in progress, which is concerned with analyzing and optimizing the creation process of MEDs. During the study, MEDs were created by physicians and evaluated by the researchers. The results of the study are pending at this time.

With regards to the technical solution in accessing MEDs on the EHC in Germany, decryption of the stored data will only be possible with an electronic health professional card. Accordingly, this restriction will limit the access to the MED to Germany only [10,11]. When individuals are traveling to foreign countries, local health professionals cannot access the data. Furthermore, only publicly insured patients will have the opportunity of storing a MED on their EHC, since the EHC is not issued to privately insured patients.

In March 2016, the three-year project Timely Information in Medical Emergencies (T.I.M.E.) started its research in the field of emergency care. The project is funded by the European Union and the Federal Ministry of Health, Equalities, Care and Aging (MGEPA) of North Rhine-Westphalia. Confronting the expected limitations of the upcoming MED implementation, one of the project objectives is to expand the previously defined concept of MEDs by emergency-relevant findings and documents. The main goal is to offer an alternative secure technical solution for a widened access to emergency-relevant information via mobile accessibility and to test and evaluate a functioning prototype to proof the concept. Additionally, a solution for the access to MEDs abroad is being investigated. Final discussions with the Commissioner for Data Protection of the federal state of North Rhine-Westphalia are currently in progress to finalize the data protection and privacy concept, which makes the architecture prone to minor changes. The goal of the current work is to describe the conceptualized software architecture.

Methods

Firstly, unstructured interviews with domain experts were conducted to determine which information should be provided with an enhanced MED and how it can be administered and retrieved. Based on these interviews use cases describing the requirements were defined. Secondly, the preexisting data model of the emergency dataset for the EHC developed by the German Medical Association and the gematik [12], the Society for Telematics Applications, which is mandated to carry out the introduction of the electronic health insurance card, was taken as a basis and an extended data model containing all relevant data from the interviews was derived. Thirdly, a concept fulfilling pre-defined use cases and the provision of an emergency dataset for patients in case of emergency was created. Finally, a data protection and privacy concept was developed and discussed with the Commissioner for Data Protection of the federal state of North Rhine-Westphalia.

Results

Use Cases

At first the requirements given by domain experts were formalized in four use cases:

- Authorizing the access to the emergency dataset by the patient
- Creating or updating an emergency dataset by an authorized physician
- Reading the emergency dataset in case of emergency in Germany
- Reading the emergency dataset in case of emergency in a foreign country

The four use cases address the central requirements, which are data security, creating and updating the MED and the procedure in case of an emergency in Germany and abroad.

Medical data model

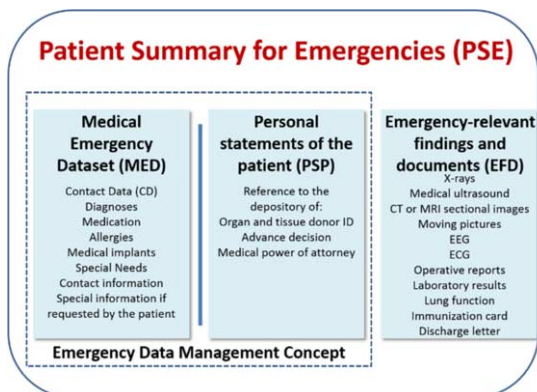


Figure 1 – Overview of the Patient Summary for Emergencies

As shown in Figure 1, based on the MED developed by the German Medical Association and gematik for the EHC, an enhanced dataset was developed. The Emergency Data Management Concept developed by gematik consisting of contact data (CD), MED and personal statements of the patient (PSP) was not altered significantly to ensure interoperability with the emergency dataset on the EHC. Additional information, for example, imaging and laboratory findings were added to the dataset under the heading ‘Emergency-relevant findings and documents (EFD)’, to provide comprehensive information for the

practicing emergency physicians. Thus, information like X-rays and physicians’ letters stored in formats like images or PDF documents can be made available for the emergency physician. In summary, all this information is stored in the Patient Summary for Emergencies (PSE).

Figure 2 shows the detailed data model of the EFD. Every finding in the EFD has corresponding finding documents and will be created by an identifiable treating physician or institution. The treating physician, who decides which information is added to the current EFD, is also documented as the administering physician of the findings in the EFD, as given in the national MED.

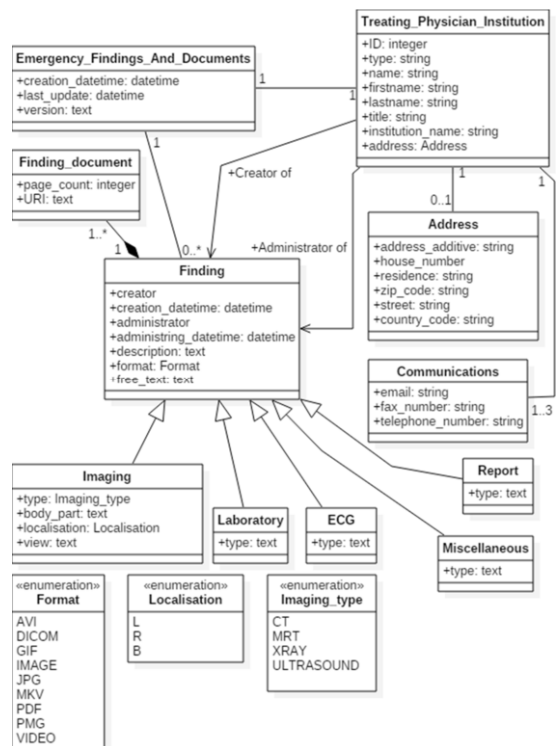


Figure 2 – The medical data model of the emergency-relevant findings and documents

Concept

CompuGroup Medical (CGM), the market leader for physician information systems in Europe, will develop the prototype. The existing infrastructure and components of CGM will be used for the implementation. Nevertheless, the concept can be generalized and developed in the same manner by other manufacturers.

As seen in Figure 3, the central component of the system is an electronic patient record called PSE, which stores the enhanced MED. All information stored in the PSE is encrypted with state of the art encryption algorithms and can only be accessed by authorized users. The authorization to access, change or add data in the record is managed by the patient him- or herself.

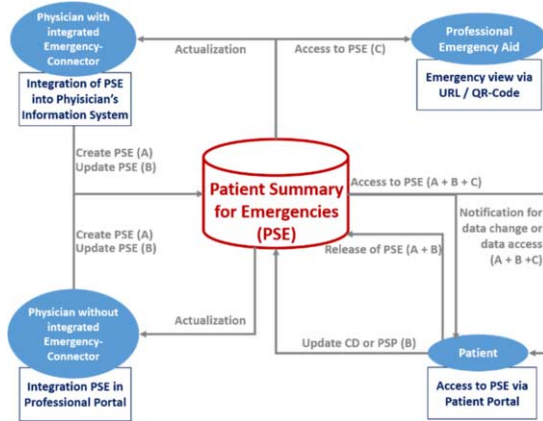


Figure 3 – Architecture overview: The physician creates (A) or updates (B) PSE for the patient (left side). The patient has to release the changed PSE (bottom right). The emergency professionals can access the PSE (C upper right).

Authorizing the access to the emergency dataset

The system is patient-centered, which means the data is stored, altered and accessed only with explicit permission given by the patient. For the creation of the PSE, the patient has to give informed consent in written form to the physician, who creates the PSE within the patient record.

Medical professionals have to be registered with their official profession within the system, like first responder or emergency physician, to access the PSE. To enable medical professionals to access the PSE, the patient can generate an access token within his/her patient portal, which is a string consisting of 25 characters. This access token can be obtained as a Quick Response Code (QR-Code), which can be opened by a preinstalled app showing the released data from the PSE. If the app is not installed an error is shown. Thus, a two-factor authentication is established. Table 1 shows the standard authorization configuration for a released PSE. The first aiders, who are defined as nonprofessional helpers like people seeing an accident and initiating an emergency call, are not granted access to the data if they scan the QR-Code. However, if a registered medical professional like the emergency service or the emergency physician uses the QR-Code, he/she can retrieve appropriate medical data regarding their needs for the emergency.

Table 1 – Authorization matrix

(R=Read, W=Write and D=Delete)

Object	CD	MED	PSP	EFD
First aider (non-professional)	None	None	None	None
First responder (professional)	R	R	None	None
Emergency service	R	R	R	None
Emergency physician	R	R	R	R
Administering physician	R,W,D	R,W,D	R,W	R,W,D
Patient	R,W,D	R,D	R,W,D	R,D

Creating or updating an emergency dataset

The creation (A in Figure 3) and update (B in Figure 3) of a PSE is done exclusively by an authorized physician, as only a treating physician who knows the patient can determine which

information is useful in case of emergency and should be part of the PSE. Only contact data and information regarding personal statements can be updated by the patient (Figure 3, bottom right corner), but he/she is not allowed to create a PSE. On the one hand, the physician can use an integrated form in his/her physician information system to manage the PSE (Figure 3, top left corner). On the other hand, he/she could use a web portal developed for physicians who do not have physician information systems or use a physician information system without PSE integration (Figure 3, bottom left corner).

Since the physician decides which information is relevant for the PSE, the patient cannot choose if specific medical information from his or her medical history or current illnesses is present in the current dataset. Thus, the PSE will always be a valid set of medical data approved by a physician. After creation or update of the PSE, the patient is notified about the changed data. Until the patient releases the new PSE or revokes his/her participation, the outdated PSE is not available in case of emergency. As the updating process is during the consultation with the physician, the time gap between an outdated and confirmed PSE should not be too long. However, the patient can revoke the permission to store the PSE at any time. If he does so or does not give his permission after an update, his complete PSE will be deleted. It is therefore an all-or-nothing approach. Since the emergency data of the patient will partly be stored on the EHC in form of the MED, the system will integrate a mechanism for updating the data on the EHC and updating the data in the PSE from the EHC.

Reading the emergency dataset in case of emergency

Registered medical professionals like the emergency service or the emergency physician can retrieve the PSE. The registration can be done by the medical professional him-/herself using the health professional card or by an authorized organization like the German Medical Association or aid organizations, who can proof the accreditation of the medical professional.

After the patient has released the PSE, he/she can print the corresponding QR-Code for the created access token. In case of emergency, the medical professional may scan this QR-Code and can retrieve the appropriate information as shown in Table 1 (C in Figure 3). The patient will be notified about every data access by other users through the patient portal and the QR-Code will automatically be disabled after a predefined period of time. Thus, the data cannot be retrieved for an unlimited time afterwards and the patient has to create a new access token to re-enable the access to the PSE. Therefore, the patient is responsible for updating the access token and carrying along the current active access token to provide access to his/her emergency dataset.

Since the expertise of health professionals in foreign countries cannot be verified, as there is no global health professional card or a standardized way to identify health professionals, the access to the PSE has to be achieved differently in foreign countries. In case of holidays or business trips, the patient has the opportunity to extend the access rights to the PSE for a specified period of time. Thus, anybody who scans the QR-Code within the specified period of time can access the whole PSE without any authorization mechanisms in place.

Data privacy and data protection

The stored information in the PSE is encrypted with state of the art industrial standard encryption algorithms. It is therefore not possible to decrypt the information with reasonable effort, without knowledge of the secret private key. Moreover, the communicated data is end-to-end encrypted and the communication between the system, which runs the PSE and the client devices is encrypted as well.

The personal and medical information of the patients are only temporarily stored on the clients' devices. After closing the app but at the latest after three hours the data will be deleted. Furthermore, the app checks if the device has been rooted or jailbroken, which leads to an error and closing the app. The access to the patients' data is managed by the patient him-/herself and the patient has the opportunity to revoke the consent on storing information in the PSE and to delete the information stored in the PSE at any time. The patient is notified every time his/her information is changed by a physician and he/she has to approve the new emergency dataset.

An audit trail for all accesses to the PSE is stored, so the patient is able to see who has accessed his/her medical information and when it happened. Furthermore, an encrypted snapshot or screenshot of the viewed data is stored within the PSE to provide the opportunity to notice abuse of information or for legal reasons.

Discussion

The objective of this work was to illustrate a software architecture for a secure technical solution to make an extended MED available to users via mobile accessibility in Germany and abroad. Built upon the current version of the official German MED model, the presented concept shows an alternative way of getting emergency-relevant information to medical professionals directly to the emergency site. The PSE, containing the Emergency Data Management components and additional emergency-relevant findings and documents, builds an enlarged basis for diagnostics and forms the foundation for this enhanced software solution. The current concept offers a way for secured access to the data via a two-factor authentication (QR-Code plus app). Only medical professionals registered within the system with their official profession, like first responder or emergency physician, are allowed to access the PSE to retrieve appropriate medical data regarding their needs for the emergency. Due to more information about the patients' medical history, the physician gains greater knowledge of patients' clinical characteristics and this is expected to improve medical decision-making [13].

To get overall clinical acceptance of the PSE, it is essential that the quality of data be guaranteed. A possible way in achieving this is if the treating physician is also the administrator of the PSE. Nevertheless, the patient could also provide additional information during the creation of the dataset since he/she can be asked specifically about emergency-relevant findings assessed by other physicians.

Concerning the limitations of this study, the PSE will exclusively be tested using the existing infrastructure and components of CGM. However, given the plans of the government for building a healthcare telematics infrastructure which includes the provision of MEDs to the public by 2018 [6], transferability of the concept to other systems has been the focus of this project from the start.

Since the PSE is primarily meant to be accessed at the site of an emergency, mobile Internet access is a vital requirement to be able to get the crucial information. Restrictive factors of mobile Internet access are of course still present in form of data volume limitations or areas with no connectivity, yet a comprehensive network covering the area of Germany is scheduled for 2018. Another fact that has to be accepted is that a mobile device, seen from the point of information security, has a residual risk of exploitation. To minimize the risk the delivered app will check the device for roots and jailbreaks. However, since emergency-relevant data should be available quickest possible, the access via mobile accessibility is crucial. Still, the patient has to be

informed about all possible security weaknesses and has to give informed consent acknowledging these data privacy concerns.

Furthermore, a privacy issue regarding accessibility of MEDs abroad might be raised, as for a defined period of time no authorization mechanism would take place when scanning the QR-code, giving access to sensitive data. This concern is a prominent discussion point in the ongoing developmental process of the concept and has also been stressed by the Commissioner for Data Protection of the federal state of North Rhine-Westphalia. While it can be argued that a patient has the sole responsibility for the safety of the QR-Code when being abroad, two alternative data protection solutions with controlled activation are currently being discussed. Firstly, a call center can grant access to the PSE to the emergency physician, after verification of the physician's profession. Secondly, implementation of a two-factor authentication sending a short message to the patient or a relative to acknowledge the access to the PSE could be an option.

Next to the limited authentication measures abroad, another issue is the translation of the content to another language. Since the dataset offers free text fields without proper semantics, automatic translation cannot be provided at this point. Furthermore, Germany has not licensed the international terminology SNOMED CT [14], which could have been used to achieve semantic interoperability. Nevertheless, the PSE will partly be mapped to international terminologies and classifications, like the International Classification of Diseases for diagnoses, Logical Observation Identifiers Names and Codes for laboratory values and Anatomical Therapeutic Chemical Classification System for medication. Additionally, there are initiatives that promote cross border exchange of personal health data like the pilot project Smart Open Services for European Patients (epSOS) [15], a six-year initiative which ended in 2014. In the project, regulatory, technical and organizational aspects were blueprinted for a cross boarder electronic health record system including the healthcare services 'electronic Prescriptions' and 'Patient Summaries'. While there are strong content-related similarities between the epSOS Patient Summary and the official German MED, the primary application area of patient summaries is in doctors' offices or hospitals. Consequently, the use of this data sharing approach is not applicable at the emergency site.

To overcome the interoperability issues, the development of an international profile, like IHE XDS, specifically addressing the emergency dataset, should be considered. Furthermore, the introduction of an international electronic health card for all citizens and an international health professional card for medical professionals would be helpful for exchanging health data and authorizing the access to health data [13].

Conclusion

In conclusion, a concept for a secure architecture to provide a medical emergency dataset for patients has been proposed. With the implementation of this architecture in its finalized state, it will be possible to supply medical professionals with emergency-relevant information concerning the current patient via mobile accessibility. Furthermore, the patient will have the opportunity to use his/her emergency dataset not only in Germany but also whilst travelling abroad.

The next activity in the T.I.M.E. project is the development of user-interfaces for the software components of the different user groups. Using these proposed user interfaces an online survey among physicians will be conducted to determine the optimal user interface for displaying MEDs. Based on the results of the survey a functioning prototype will be developed, tested and

evaluated. Finally, a support infrastructure (e.g. telephone hotline) has to be constructed for the test phase and possible future commissioning.

Acknowledgements

This work is supported by and the European Union and the Ministry of Health, Equalities, Care and Ageing of the State of North Rhine-Westphalia (MGEPA <http://www.mgepa.nrw.de>) Grant ID: GE-1-1-006A / EFRE-0800196. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- [1] J.S. Shapiro, J. Kannry, A.W. Kushniruk, and G. Kuperman, Emergency physicians' perceptions of health information exchange, *J Am Med Inform Assoc* **14** (2007), 700-705.
- [2] A. Stiell, A.J. Forster, I.G. Stiell, and C. van Walraven, Prevalence of information gaps in the emergency department and the effect on patient outcomes, *CMAJ* **169** (2003), 1023-1028.
- [3] J.T. Finnell and J.M. Overhage, Emergency medical services: the frontier in health information exchange, *AMIA Annu Symp Proc* **2010** (2010), 222-226.
- [4] N. Stafford, Germany is set to introduce e-health cards by 2018, *Bmj* **350** (2015), h2991.
- [5] gematik, For a health-care system with a future. The electronic health insurance card, [cited 12. December. 2016], Available from: https://www.gematik.de/cms/media/infomaterialpresse/_Broschuere_Englisch_2.pdf
- [6] Federal Ministry of Health of Germany, Act on secure digital communication and applications in the health care system (E-Health Act), [cited 15. December. 2016], Available from: http://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&jumpTo=bgbl115s2408.pdf
- [7] J. Born, J. Albert, A. Bohn, N. Butz, K. Fuchs, S. Loos, J. Schenkel, and C. Juhra, Der Notfalldatensatz für die elektronische Gesundheitskarte, *Notfall+ Rettungsmedizin* (2016), 1-6.
- [8] German Medical Association, [cited 12. December. 2016], Available from: <http://www.bundesaerztekammer.de/weitere-sprachen/english/german-medical-association/>
- [9] J. Born, J. Albert, N. Butz, S. Loos, J. Schenkel, C. Gipp, and C. Juhra, The Emergency Data Set for the German Electronic Health Card-Which Benefits Can Be Expected?, *EHealth2015-Health Informatics Meets EHealth: Innovative Health Perspectives: Personalized Health* **212** (2015), 206.
- [10] B. Blobel, Comparing approaches for advanced e-health security infrastructures, *Int J Med Inform* **76** (2007), 454-459.
- [11] N. Ernstmann, O. Ommen, M. Neumann, A. Hammer, R. Voltz, and H. Pfaff, Primary care physician's attitude towards the German e-health card project--determinants and implications, *J Med Syst* **33** (2009), 181-188.
- [12] gematik, [cited 12. December. 2016], Available from: <https://www.gematik.de/cms/de/startseite/index.jsp>
- [13] M. Moharra, C. Almanán, M. Decool, A.-L. Nilsson, N. Allegretti, and M. Seven, Implementation of a cross-boarder health service: physician and pharmacists' opinions from the epSOS project, *Family practice* (2014), cmv052
- [14] International Health Terminology Standards Development Organisation (IHTSDO), Snomed CT, [cited 12. December. 2016], Available from: <http://www.ihstso.org/snomed-ct>
- [15] epSOS – the European eHealth Project, [cited 12. December. 2016], Available from: <http://www.epsos.eu/>

Address for correspondence

Michael Storck, michael.storck@uni-muenster.de

An Architecture for Semantically Interoperable Electronic Health Records

André Toffanello^a, Ricardo Gonçalves^a, Adriana Kitajima^a, Ricardo Puttini^a, Atualpa Aguiar^b

^a Core Consulting, Brasília, Distrito Federal, Brazil

^b Hospital Alemão Oswaldo Cruz, São Paulo, São Paulo, Brazil

Abstract

Despite the increasing adhesion of electronic health records, the challenge of semantic interoperability remains unsolved. The fact that different parties can exchange messages does not mean they can understand the underlying clinical meaning, therefore, it cannot be assumed or treated as a requirement. This work introduces an architecture designed to achieve semantic interoperability, in a way which organizations that follow different policies may still share medical information through a common infrastructure comparable to an ecosystem, whose organisms are exemplified within the Brazilian scenario. Nonetheless, the proposed approach describes a service-oriented design with modules adaptable to different contexts. We also discuss the establishment of an enterprise service bus to mediate a health infrastructure defined on top of international standards, such as openEHR and IHE. Moreover, we argue that, in order to achieve truly semantic interoperability in a wide sense, a proper profile must be published and maintained.

Keywords:

Electronic Health Records; Health Information Exchange; Vocabulary, Controlled.

Introduction

Different systems are said to interoperate when they can provide an expected functionality on top of a common interface. Narrowing the context to electronic health records (EHRs), a common way to label interoperability relates to its depth [10]: syntactic, functional or semantic, according to whether the message is properly parsed by machines, humans or both. Keeping that concept in mind, this work tackles the current lack of interoperability, a major barrier in current medical systems.

There are a few suggested approaches in the known literature. Ryan and Eklund [6] coined, several years ago, the idea of the Health Service Bus (HSB), acting as a middleware that routes and transforms services, in order to mediate interactions with a common infrastructure. However, the described approach is cumbersome and hard to maintain. We argue that, instead of a translation approach, it is more suitable to build a future-proof architecture from the ground up.

Despite the availability of some quite respectable references when it comes to establishing a wide e-health strategy [9], if we take a look at the Brazilian scenario, it is easy to identify a silo architecture, inherited due to the compatibility with legacy systems. At this point, not only are the services coupled, but the implementation is also tied to the data persistence mechanism. This leads to a problem-oriented architecture that makes semantic interoperability practically unreachable.

On the other hand, replacing an architecture implies great effort and requires strong reasons. Current scenarios usually focus mostly on administrative and financial data, providing poor clinical value. This is a problem that also needs attention. After all, if we are looking to a more appropriate solution, this is also the opportunity to improve the quality of healthcare.

This work is driven towards a reference architecture, wherein it is extremely important to prioritize the knowledge artifacts. Nonetheless, in order to present a future-proof solution, a mature and self-contained theoretical proposal is desired. This is why our work is built on top of international standards such as IHE [2] and openEHR [1]. Although this combination is not new to the literature [5; 7], the known texts lack functions vital to semantic interoperability, such as mapping terminologies [4]. Other aspects, such as identity management, consent, security and privacy also apply, but are out of the scope of this paper, even though they share the same principles summarized here with a common goal: to facilitate the access to healthcare data.

This project is supported by the Hospital Alemão Oswaldo Cruz through the PROADI-SUS program, a governmental initiative that encourages excellent hospitals to propose and execute projects within topics defined by the Ministry of Health. Moreover, it is quite beneficial that we can implement this design in a controlled region, while still interoperating with other parties in a way that it can coexist or expand nationally.

The next sections are organized to first introduce our methods, which are the standards and techniques orchestrated to compose the solution, followed by the resulting architecture, a discussion of the contribution to the current scenario and the conclusions.

Methods

This section details the building blocks of the architecture, built on top of international standards. The main goal we introduce is to achieve EHRs that both machines and humans understand. As for the first part, the IHE standards provide a quite solid specification for distinct health organizations to exchange data, tackling syntactic interoperability. When it comes to semantics, even though solutions such as openEHR already take a step further in bringing meaning into messages, there are pitfalls, e.g. terminology mapping and management of clinical artifacts.

IHE

When we defined interoperability, one of the key points was to be understood by a machine. This is exactly where Integrating the Healthcare Enterprise (IHE) steps in, introducing an IT infrastructure framework described by a set

of profiles that standardize all the defined transactions between the actors, including the messages exchanged.

Profiles

As stated, IHE standardizes transactions in a way that they can be processed by the machines. The guidelines are published and maintained within modules known as profiles.

There are various profiles, but a few ones are listed below:

- ATNA (Audit Trail and Node Authentication): security measures, data integrity and accountability;
- CT (Consistent Time): time synchrony between multiple actors and computers;
- XCA (Cross-Community Access): query and retrieve medical data within other communities;
- XUA (Cross-Enterprise User Assertion): communicate identity claims in transactions that cross affinity domain boundaries.

There are also three more profiles (PIX, PDQ and XDS.b) highly important to this discussion. They will be discussed next.

PIX

The Patient Identifier Cross-Referencing (PIX) profile provides a mechanism to correlate patient identifiers from different domains in a common module.

It provides not only querying and retrieval of identifiers, but also a feed notification, so that other actors can react immediately to related events.

PDQ

Somewhat related to PIX, the Patient Demographics Query (PDQ) profile allows an actor to resolve identities based on demographic information.

It also allows the system to customize the level of exposure, controlling the amount of demographic information shown and the associated input filters.

XDS.b

The Cross-Enterprise Document Sharing (XDS.b) is the most relevant profile when it comes to syntactic interoperability and describes two important actors within our architecture: the document repository and the metadata registry. It is based on the OASIS ebXML specification, the *de facto* standard for interoperability of content through standardized transactions.

The document repository acts as a distributed document database, serving the exchanged documents.

The metadata registry is unique. It catalogs a set of metadata for each document available across the system, indexing submissions coming from different repositories. The community established around a registry is called an “affinity domain” in the IHE vocabulary.

Every interaction within an IHE actor happens through a well-defined transaction, commonly denoted by ITI-*n*. For example, ITI-41, ITI-42, ITI-18 and ITI-43 refer to submitting, registering, querying and retrieving documents, respectively. More information is available in the referenced documentation.

It is important to notice that the XDS.b is an integration profile, completely neutral when it comes to the document contents, which is why we say it contributes to syntactic interoperability.

openEHR

Taking a step further into the goal of semantic interoperability, openEHR is defined as a domain-driven platform for e-health.

It is actually a dual-layer approach that introduces a new way of modeling knowledge artifacts.

Beale et al. [1] argue that interoperability is difficult to achieve because common single-model systems use different models. This is how openEHR stands out.

Reference model

The bottom layer is the reference model, a set of stable specifications at the core of the data layer that provides a finite number of data types and structures to be processed by the platform during compilation time.

Archetypes

Upon the bottom layer, there is a top one that deals with organic units of information, usually known as archetypes, at runtime. The advantage of such an approach is that archetypes can be maintained by domain specialists, without the need to deal with complicated languages or technicalities from the bottom layer, following the principle of separation of concerns.

We can think of an archetype as a serialization of a structured information tree, wherein the upper nodes are use cases customizable according to specific business needs and the leaf nodes are classes formally defined within the reference model.

Templates

As defined by openEHR, an archetype is supposed to be an organic unit of clinical information, such as blood pressure, to be reused in different contexts.

In order to achieve something closer to a clinical encounter, openEHR allows compositions, normally known as templates, to be outlined as a combination of archetypes with equivalent or narrower restrictions related.

From a computer science perspective, this actually states that an archetype in a template is a subclass of its parent definition, stating an inheritance relationship wherein the child can be seen as a valid instance of the parent, *e.g.*, backward compatibility.

Revisiting our tree analogy, the template is the root and each of its subtrees configures an archetype developed and used locally.

CTS2

ISO ANSI [3] clarifies that semantic interoperability is not binary. There are degrees depending on the level of agreement on ontology, which is why in our work the terminology server is independent and self-centered.

The Common Terminology Services 2 (CTS2) is a specification resulting from extensive work by both HL7 and OMG. On top of authoring and serving code systems and value sets, the key definition is that the terminology server is able to translate concepts on machine and human domains, *e.g.* code and terms.

We also have to take into account that different health professionals may have distinct ontologies in their own domain. It is expected that such associations can be mapped and reproduced as knowledge, regardless of how the data is stored, so that it is also possible to have different terminologies when creating and retrieving a document, depending on local needs.

SNOMED CT

When we bring to the table the mechanistic linking of ontologies, it is common to question the weight of a single terminology. However, a reference is still needed, so that all the mappings refer from or to this major set of clinical terms.

SNOMED CT [8] is the most comprehensive and precise clinical health terminology product that we are aware of. Therefore, it is the elected reference ontology, so that the community itself can act on mapping a local terminology in a way that it becomes available and understandable for others. This can be done directly in the CTS2 module, wherein such mappings go through a formal governance, or more loosely on an openEHR document, wherein local mappings can be documented and later collected as an incoming data set to be processed by the backend itself.

Results

In this section, we demonstrate the consolidated architecture. First, we start with a general overview of the Brazilian HSB, then we highlight several interesting aspects.

The Brazilian Health Service Bus

The core proposal of this work is the design shown in Figure 1. We propose an Enterprise Service Bus (ESB) to serve vendor-specific implementations behind the standard contracts, defined by IHE. The document repository and metadata registry interoperate with an existing PIX/PDQ layer, which is, in turn, connected to a Master Patient Index (MPI) infrastructure. So far, everything conforms to the IHE ITI Technical Framework. But now there are important additions: the CTS2 server and the Clinical Knowledge Manager (CKM) module, both exposed as services. They are expected to provide ontology mapping and knowledge artifacts, respectively.

While it is expected from the environment to somehow interoperate with the CTS2 and CKM services, such communication and governance schemes are not standardized. The resulting ecosystem relates not only the known IHE and openEHR organisms, but also actors related to knowledge artifacts governance, terminology mapping and access points. Therefore, although the document exchange and modeling are specified, this architecture also tackles the remaining gaps that prevent heterogeneous entities from sharing the infrastructure.

Semantic interoperability

ISO ANSI [3] identifies four prerequisites for semantic interoperability:

1. Standardized EHR reference model: comes from the openEHR reference model;
2. Standardized service interface models: while there is an ambition regarding the openEHR service model, currently in development, it is not wide enough for semantic interoperability. This work introduces an overview of the functional architecture, but the service interface models must be defined based on local needs;
3. Standardized set of domain-specific concept models: the openEHR archetypes and templates;
4. Standardized terminologies: SNOMED-CT is taken as the reference; archetypes and templates are free to bind according to specific needs, as long as the related mappings are available, relying on the CTS2 for that.

While we have discussed all of them, the practical contributions from this paper relate mostly to the composition of a standard service interface model and ontology resolution.

Improvement of health care

Not only is the proposed architecture built on top of recognized international standards, but it is also designed to be patient-centric. It is ready to facilitate access to healthcare data and also to take part in even broader e-health integration scenarios.

Integration between health processes

With the siloed architecture gone, we are ready to embrace a service-oriented architecture. All transactions drawn from IHE are inherited from ebXML, a mature set of enterprise standards.

Decision support

As soon as we reach the capability of semantic interoperability, it is finally reasonable to think about clinical decision support. Now that systems and people are able to exchange and understand the underlying clinical information,

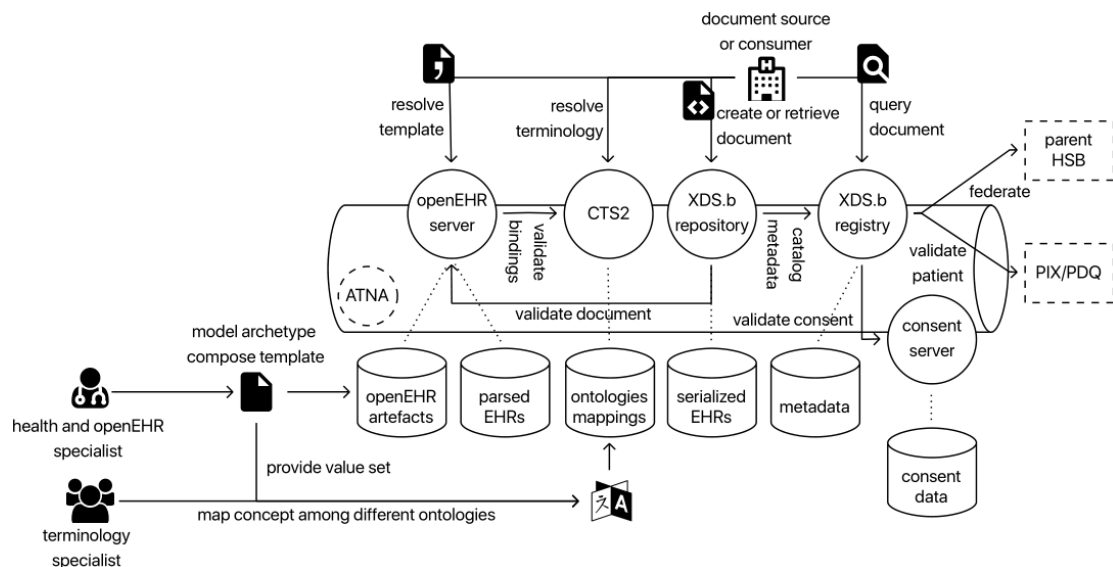


Figure 1—Functional architecture of the Brazilian Health Service Bus

practitioners can consume all healthcare data like a timeline, in such a way that it sets a full-featured EHR.

Epidemiology and public health

Considering that demographics and healthcare data are stored separately, it is now possible to compute statistics and population metrics. For example, by using AQL queries, we can select records based on contents and even reshape the document structure according to specific needs (e.g. big data and BI). Moreover, we can also idealize custom triggers that monitor and track the behavior of specific parameters within an observation range and fire customized alerts for health events.

Federation

Even though IHE states that the document registry must be unique within an affinity domain, we find it mandatory for distinct communities to be able to interoperate in a wider sense than specified by the XCA profile. Actually, our architecture is designed so that HSBs can be established as siblings (e.g. independent regional deployments aware of each other) while federated under a parent and eventually its ancestors (e.g. national deployment that catalogs all available health records). In other words, this work elaborates an HSB mature on its own but ready to coexist and federate.

Discussion

At this point of the document, we introduce new discussions that unfold from the current results.

Service model and openEHR REST API

With the openEHR clinical modeling features proven valuable, a lot of work from the openEHR Foundation emerged regarding a standardized REST API to handle knowledge artifacts. While it is likely not to replace the IHE layer, as such an approach to exchange data among parties is already in use by the industry, it can surely provide a more lightweight service interface that refines the granularity of the EHR to more specific levels, such as templates, archetypes and even data, instead of the document-oriented style from IHE.

With such developments in mind, it makes sense to have a dedicated openEHR server within the architecture, so that we have more than just documents, such as the Archetype Query Language (AQL), which enables us to operate on the models, instead of the serialized data itself.

On the other hand, as the openEHR REST API delivers functionality beyond the capabilities of the IHE specification, we already have an orchestration infrastructure to enable controlled usage, either internally or externally.

openEHR and FHIR

Fast Healthcare Interoperability Resources (FHIR) is a recent standard proposed by HL7 as an API to enable parties to exchange data, aiming for less complexity and more flexibility, especially when compared to HL7v2 and HL7v3 messages. It works upon the idea of resources and profiles, *i.e.* reference model and archetypes/templates within the openEHR scenario. Although not exactly the same, it is possible to translate them, so we can embrace an even more heterogeneous environment, and such an implementation design is suggested for parallel work.

Management of knowledge artifacts

Along with all the orchestration we have presented so far there is the assumption that the knowledge artifacts are reachable. This can get in the way of semantic interoperability, especially

when different vendors coexist within the same infrastructure, because there is no standard way to manage archetypes and templates. The same goes for validating external terminologies.

That being said, there are several approaches going on, most lead by implementing vendors. Despite having, for example, a CTS2 standard to interoperate with a terminology server, we lack a formal definition of how an EHR would consume it. Additionally, within the openEHR context, there is the Clinical Knowledge Manager (CKM), a distributed portal to handle archetypes and templates encompassing a reasonable life cycle, including approval, translation and versioning of objects. However, again, there is no standardized way for one actor to interoperate with such functionality.

Interoperability with different affinity domains

Most of the solutions known so far assume a set of document repositories working together under the umbrella of a single metadata registry, which configures an affinity domain. On the other hand, the architecture we propose satisfies the possibility of multiple affinity domains coexisting in a federated fashion. Along with this scenario, new concerns about security and privacy emerge, demanding a more prolific governance profile.

Future work

Even though this architecture achieves the goal of exchanging understandable health data, there are important use cases other than immediate care that could be practical for further work. We discuss several topics below.

Case management

As soon as we have a good healthcare data architecture with basic semantics aiming for decision support, the spotlight moves towards workflow support.

The objective is to come up with a collaborative model capable of covering a whole episode, *e.g.* ongoing care and monitoring, different professionals and the social scenario of the patient. Besides the need to systematize medicine within protocols, it is also needed to formalize such concepts in a computable way that handles decision points with logic. Even at this point, as much as we translate care for conditions into protocols, care for a person may follow infinite paths because of many factors.

The openEHR bottom-up approach somewhat tackles decision support, starting with an extensible standard state machine to act like a GPS system, even capable of correcting wrong turns.

Internet of things (IoT)

Through the progress of the IoT, home and self-care devices will become popular quite quickly. However, there are several difficulties to be clarified. Above everything, we need a secure interface for those devices to aggregate data.

Not only is such data highly transactional, meaning it could flood an EHR with irrelevant or repetitive information, but also the responsibility for its accuracy is not clear.

Secondary use

Beyond the boundaries of primary care, secondary use implies extracting information not about the patient itself, but about the healthcare ecosystem instead.

Now that we have meaningful data in a structured way and detached from any demographic information, the idea of computing population statistics and triggering disease control programmatically becomes perfectly feasible, as long as

maintainable policies regarding usage consent are available. The same applies for management and research.

It is important to acknowledge that the openEHR approach for querying and shaping EHRs as needed with AQL allows the processing of clinical data on a functional level of granularity, e.g. an archetype, regardless of the templates used to input data, relying on the object model instead of the serialized message. Moreover, range comparison, filtering and data sorting functions included with the querying language enable governments to compute over heterogeneous and huge datasets, feeding big data and business intelligence systems accordingly.

Conclusions

When it comes to health data, semantic interoperability cannot be achieved if systems using disjoint standards don't cooperate.

Now, taking a step back in the discussion conducted so far, we built a complex architecture on top of the known IHE ecosystem, tackling a more complex situation wherein a federated infrastructure is expected to validate openEHR documents with bindings to external terminologies.

This work also elaborates on the management of knowledge artifacts, more specifically debating the usage of the CKM to handle archetypes and templates within a reasonable life cycle that makes sense to domain specialists.

At last, we pointed out the recent work on the openEHR REST API and introduced a possible compatibility path to expand our environment towards FHIR systems.

A new profile

A major observation presented here is the desire for further standardization in order to achieve semantic interoperability. The proposed architecture is a suitable starting point, wherein a reference organization such as IHE can elaborate in order to develop a more solid specification.

Several important requirements can be drawn from the gaps that our mediation is intended to overcome, namely:

1. Provide semantic meaning to the shared documents: IHE exchanges binary data with a MIME type and some arbitrary classification metadata, which is not enough for it to be processed as a true document by the recipient.
2. Allow consumers to map local ontologies as needed, reducing the impact for different entities to collaborate.
3. Allow consumers to locate and collect models and resources published by others that are relevant to handle instances of such data (e.g. archetypes and templates).
4. Specify a core set of features on top of the model itself, in order to prevent vendor lock-in.
5. Provide a standard for patients to manage consent and trace it along with their documents. Such information must be available within the infrastructure as well so that the policies can be enforced to exchange health data.

By satisfying the criteria described so far in this work, upcoming specifications are closer to achieving a semantically interoperable electronic health record wide enough to support an entire country.

Acknowledgements

We would like to thank the Hospital Alemão Oswaldo Cruz for the support through the PROADI-SUS program.

References

- [1] T. Beale, S. Heard, D. Kalra, and D. Lloyd, OpenEHR architecture overview, *The OpenEHR Foundation* (2006).
- [2] Integrating the Healthcare Enterprise, IT Infrastructure Technical Framework, *Cross-Enterprise Clinical Documents Sharing (XDS)*, by the IHE ITI Technical Committee, Version 3 (2006).
- [3] ISO ANSI, TR 20514 Health informatics—Electronic health record—Definition, scope and context, in, ISO, 2005.
- [4] C. Martínez-Costa and S. Schulz, Ontology content patterns as bridge for the semantic representation of clinical information, *Applied clinical informatics* 5 (2014), 660.
- [5] C. Pahl, M. Zare, M. Nilashi, M.A. de Faria Borges, D. Weingaertner, V. Detschew, E. Supriyanto, and O. Ibrahim, Role of OpenEHR as an open source solution for the regional modelling of patient data in obstetrics, *Journal of biomedical informatics* 55 (2015), 174-187.
- [6] A. Ryan and P.W. Eklund, The Health Service Bus: an architecture and case study in achieving interoperability in healthcare, (2010).
- [7] M.R.d. Santos, Sistema de registro eletrônico de saúde baseado na norma ISO 13606: aplicações na Secretaria de Estado de Saúde de Minas Gerais, *Perspectivas em Ciência da Informação* 16 (2011), 272-272.
- [8] SNOMED CT, Systematized nomenclature of medicine-clinical terms, *International Health Terminology Standards Development Organisation* (2011).
- [9] World Health Organization, *National eHealth strategy toolkit*, International Telecommunication Union, 2012.
- [10] R. Ye, Semantic interoperability in electronic health record: a standardised approach, (2015).

Address for correspondence

André Toffanello

SIG, quadra 01, lote 385,

Platinum Office, sala 336

70610-480

Federal District, Brasília, Brazil

E-mail: andre.toffanello@coreconsulting.com.br

A Healthy Lifestyle Intervention Application

Trond Tufte^a, Ankica Babic^{a, b}

^aDepartment of Information Science, University of Bergen, Bergen, Hordaland, Norway

^bDepartment of Biomedical Engineering, Linköping University, Linköping, Sweden

Abstract

In this project, a mHealth tool for smart-phones has been developed using Design Science methodology, where the goal has been to promote an active lifestyle. This was undertaken by implementing social and physical activity stimulating features within the application MoveFit. Users can opt to utilize just a feature or two or engage in social activities of different intensity. Regular and expert users have evaluated the application in order to meet usability requirements. In addition a field expert and a focus group have contributed towards the application's potential to increase physical activity.

There was enough data collected by the app to document its good effect; it was possible to demonstrate that the app was capable of promoting physical activity. User testing has also shown the appreciation of the various features such as social networking, activity monitoring, and route/activity creation.

Keywords:

Sedentary Lifestyle; Application; Social Networking

Introduction

According to World Health Organization the worldwide prevalence of obesity nearly doubled between 1980 and 2008, emphasizing its level of significance [1]. Obesity is often resulting from a sedentary lifestyle, which is also often connected to depression, cardiovascular diseases, diabetes, and some types of cancer. Sedentary lifestyle seems to have become a quite common problem, some people tend to be less active, which in conjunction with unhealthy eating habits, is resulting in high rates of obesity [1]. Technologies such as the Computer and TV are influencing people to stay inactive for longer periods of time.

The World health organization (WHO) has stated that inactivity is the fourth most leading risk factor for global mortality. What is surprising is the severity of health risks from this sedentary lifestyle. Harvard Medical School finds an accelerated risk of diabetes, high cholesterol and heart disease, primarily due to cardiovascular problems.

It is interesting to note that this applies to the “average” person who now spends more than half their waking day sitting still. The American Journal of Preventative Medicine [2] echoes these findings, also concluding that they cannot simply be dealt with by occasional strenuous exercise like going to the gym once in a while. The rapid technological advances which have led to this problem e.g., computers in the office, can also be re-applied toward designing a non-intrusive solution, that will address this problem. While it is often necessary to sit still for

hours on end, there is no need for this time to be uninterrupted. For the rest of our careers, good habits in terms of taking regular breaks will prove invaluable, and ways to emphasize this in the workplace, schools or at home is essential.

mHealth (mobile health) is an emerging field that addresses issues such as sedentary lifestyle by use of mobile technologies. The project presented here has focused on developing a mHealth app that counters sedentary lifestyle, by allowing the users to create routes/activities, social networking, and activity monitoring. The application is location based i.e., the users find activities/routes or other users in the nearby area.

MoveFit Application

There were several research questions to be answered prior to and through the development of the application. Some of those are: ‘Is it possible for a mHealth application such as MoveFit to reduce sedentary lifestyle?’ and ‘Can a mHealth application such as MoveFit promote psychological health?’

The design of the app includes features to answer those questions and to address various user groups' needs. This allows various degrees of engagement and activities tailored to suit the different user groups.

Some of those potential users need to be just initialized into a more active lifestyle while the others would just need to be challenged more to diversify their activities and use social media to meet in person, as well as getting more active and social at the same time. This is one of the major challenges for the social media user in general. Furthermore it is possible for people all ages to do or create activities that are corresponding to their respective activity level. For instance a relatively inactive person can start off with an easy activity, with a shorter distance and from there on progress gradually.

Methods

This project has utilized Design Science and System Development as a multi-methodology. Hevner's 7 guidelines were used to design the artefact [3]. Several system development methods were used. Prototyping was the suitable model, whereas TDD, XP and Personal Kanban were methods and frameworks applied. Both qualitative and quantitative methods were used for empirical research, resulting in mixed methodology usage.

The application was evaluated by both end users and expert users. These two groups were observed while assessing the system. A focus group was also participating in the system evaluation, from the first prototype to the finished product. Furthermore a field expert was consulted in order to shed light

on the effect this application can have to the mental issues related to sedentary lifestyle. This will be elaborated more in the Results Chapter.

Quantitative Survey

For the quantitative research three different surveys have been used. The surveys have been both structured and semi-structured in the sense that they have predefined response alternatives, as well as comment fields. The first survey is for “normal” users i.e., end users without any required technical background in IT. This survey is based on System Usability Scale (SUS) [4] and it aims to find out how non-experts handle an artifact such as the app developed in this project. The intention with that was to see whether it is easy or hard to use and understand the app. The users were given predefined tasks to perform before answering the survey.

The second survey was aimed at “expert” users i.e., those who have background in IT. They were given the same set of tasks to perform and then answered the survey Nielsen's heuristics [5]. As with the first survey, it was structured.

The last survey was more related to the actual users, in terms of what their subjective opinions on the application are and physical activity. This survey was semi-structured as it is possible to write comments in addition to the predefined answers. This survey was given to the focus group.

Qualitative Interview

The qualitative research is in this project based on interviews and observations. This type of interview was an expert elicitation interview. Formal expert elicitation is a structured approach to systematically consult experts on uncertain issues. It is often used where there is a lack of data or knowledge on a topic and to further develop qualitative issues such as definitions or assumptions [6]. In this case the expert was interviewed to get a better understanding of how physical activity and mHealth can affect mental health. A licensed psychologist working at Helse Bergen, with three years of counselling experience, was elicited to provide an expert insight. The particular goal was to get an opinion on whether the developed artefact actually had the potential to help answer the research questions at hand.

Qualitative focus group

A focus group is a form of group interview where there are several participants in addition to the moderator.

There is also an emphasis in the questioning on a particular fairly tightly defined topic [7]. However, in this case, the focus group was a bit different than the conventional focus group, as it has operated via social media, i.e., Facebook, Skype, and email, and can therefore be considered a form of social media focus group [8]. The members did not meet physically in focus group session, but they were aware of each others presence in the prototype design and implementation and could communicate via messages in the app. The members of the focus group have contributed from the initial development phase to the very end. They have served as advisers in the sense that they have tested the application as a prototype for each iteration and provided suggestions for both improvements and new content. The focus group have consisted of five people who were all healthy; the age ranged from 25 to 57 years. They were motivated to improve their sedentary lifestyle and ideally prevent chronic diseases.

Qualitative Observation

Observation is another qualitative method that has been used in this project during the evaluation and data collection process. In this case controlled environment observation has been used which refers to the participants who are performing specific tasks in a controlled environment [9]. While the participants were performing the given tasks, the evaluation leader was observing them to see what kind of problems that might occur. This is useful as the evaluators can communicate their problems as they happen and the evaluator leader can take notes and analyze them later.

Results

This chapter will give an overview of the system evaluation results, as well as the findings from the data collected by the app.

End User Evaluation Results

In the figure below (Table 1) we can see the results from the end user evaluation. There were 10 evaluators. They were given 10 predefined tasks to execute before they were given a survey. They were given the well-known System Usability Scale (SUS) survey.

Table 1 – System Usability Scale Scores

User	Q 1	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8	Q 9	Q10	Score
1	1	2	2	1	1	5	2	5	3	1	37.5
2	2	1	4	1	4	2	5	3	3	2	72.5
3	1	3	2	3	3	4	5	5	1	2	37.5
4	1	0	1	1	1	5	5	0	1	1	Incom
5	1	1	5	1	5	1	5	1	5	5	80
6	4	1	5	1	5	1	5	1	5	1	97.5
7	2	2	4	1	3	2	5	4	4	1	70
8	4	2	4	1	4	1	4	3	4	1	80
9	5	1	5	1	5	1	5	1	5	1	100
10	4	1	5	1	5	1	5	1	5	1	97.5

The average score from the evaluation was 74.7 / 100 which is considered a good score. One of the evaluators did not complete the survey and was therefore excluded from the final score.

Expert User Evaluation Results

As with the end user evaluation, there were 10 users performing the 10 same tasks. They were given a survey after completion. The evaluators were given Nielsen's Heuristics to assess the system. The figure below is showing an answer of

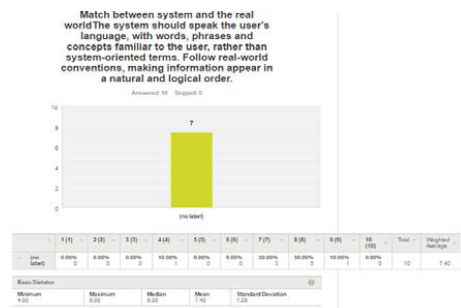


Figure 1– Nielsen's Heuristic Sample.

one of the heuristics with corresponding statistics. As we can see, this particular aspect scored 74.4 / 100.

“Documentation and Help” heuristic for instance, only scored 4.6 and ought to be improved in future development iterations. The total score from the evaluation was 6.1 / 10. Indicating that there is some room for improvement in terms of user-interface and usability of the app.

Focus Group Evaluation

After publishing the application the focus group was asked to give their assessment. They were given a semi-structured survey. One of the questions given was “Which features of the application do you like the most?”. We can see from the figure below that the route/activity creation and the social features were the most popular ones.

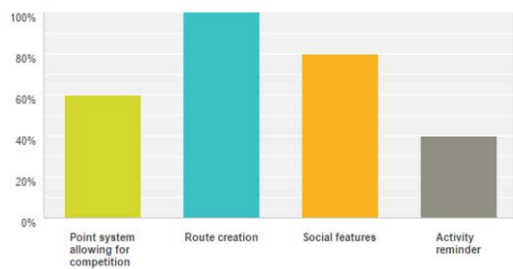


Figure 2– Popular Features

Field Expert Elicitation

A psychiatrist from Helse Bergen was interviewed in regards to the app's potential to effect mental issues related to sedentary lifestyle.

Interview Summary

The interviewee shed light on several important questions. For instance the fact that sedentary lifestyle indeed does negatively affect mental health and in those cases physical activity is recommended to counter sedentary lifestyle. It was also brought forth that the health institution Helse Bergen does not enforce use of mobile technologies as a part of patient treatment, although there are apps that are recommended for that purpose. The interviewee seemed rather unsure whether for instance hospitals ought to use mHealth apps which would depend on the patient and the condition. She would prefer to use dialogue instead. The interviewee was also positive when asked if moveFit could have a good effect on physical and mental health, with the exception of people who are really struggling. After being asked about any improvements, she suggested several interesting features that could be implemented. For instance a patient tracking system, where the therapist could communicate with the patient and follow up the patient's progress.

Application Data

Since the app was published on Google Play in late September of 2016, the application has been collecting user data. This data is related to the application core features. The core features of the app are:

- Route/Activity Creation
- Activity Alarm
- Social Networking

The user can create their own routes or activities and earn points depending on the route type and length. These points are used to compete with other people in the area. Figure 4 below is showing an example of a route create by a user.

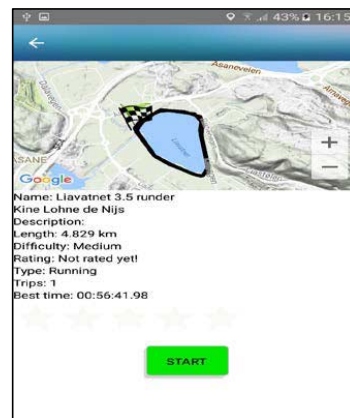


Figure 3– Sample Route

The activity points are visible in a scoreboard. There is one scoreboard for the user and one for the routes. Routes are sorted by highest review, while points sort users. Figure 4 is illustrating this.

Figure 4– Scoreboard Users

Scoreboard People			
Name	Age	Gender	Score
[Profile Picture]	32	Female	1998
[Profile Picture]	56	Male	1350
Trond Kjetil Tufte	27	Male	1231
[Profile Picture]	26	Male	733
[Profile Picture]	26	Male	100
[Profile Picture]	29	Female	0
[Profile Picture]	31	Female	0
[Profile Picture]	25	Male	0
[Profile Picture]	27	Male	0

Discussion

The Design Science approach was utilized in this research as the guiding principle to answer the research question. Following the guidelines by Hevner et al. [10] for doing design science research- was very helpful for securing the holistic structure of the research conduct. The final results of the research, following these guidelines are: a functional application (app) developed and evaluated using well defined methods. These methods regard the feasibility of the app to promote physical activity and to utilize the ways that the social medial functions. There is another social dimension: besides aiming at developing a potential to improve a general well-being and health at a personal level, the society at large could in turn benefit through reducing health costs and improving the health care outcomes. The Design Science framework enables another step and that would be testing the application in a clinically controlled manner such as a clinical trial.

In this project a mobile application has been developed in order to see to which extent it can influence health by encouraging

the users to be more active. The goal is to counter a sedentary lifestyle. The developed app is offering different functionalities that promote activities, which was also evaluated by 25 users. The evaluation was positive and constructive, the focus group reported increased physical activity. Furthermore, potential of the app to improve sedentary lifestyle was appraised by the field expert. The data collected from the application is showing that users have been physical active. They have created/completed routes or activities, established friendships, and responded positively to the activity alarm, indicating that the application has had the intended effect. The amount of active users is relatively low, and therefore a larger test should be conducted over a longer period, with more users, to get a definite answer to which extent physically activity is increased. *Figure 5* below shows the activity alarm screen.

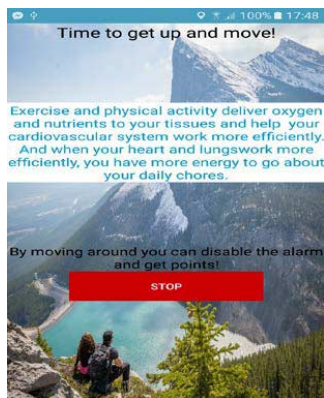


Figure 5– Activity Alarm

Other research has also shown that mobile health technologies have had a positive effect on physical health. Actual change in health this far is hard to determine as it has not been tested with clinical endpoints, although there is research connecting physical activity to better health in general.

From the experts point of view it would be possible to promote psychological health with mHealth applications and more specifically with the application presented here. The main reason behind this is due to the social and activity promoting features contained in the app. Although this is the case in general, it was brought to attention by the expert that people suffering from for instance severe depression, would most likely not use these types of applications in the first place. Furthermore, the expert expressed that the ideal situation would be to use this kind of application in conjunction with conventional therapy, such as talk therapy. Previous research has shown positive results in terms of promoting psychological health with mHealth. Fox [11] has shown that psychological health is positively affected by physical activity and that exercise is used as treatment on e.g., clinical depression.

Conclusion

This research addressed a crucial problem in our society, namely sedentary lifestyle and chronic diseases caused by it. The main scientific approach that was used is Design Science methodology. The resulting artifact is a social fitness app. The application appeared to have a positive effect on the users as well as the focus group in terms of promoting physical activity. User testing has also shown the appreciation of the various features such as social networking, and route/activity creation. The system evaluation has reached satisfactory scores

in terms of functionality and usability. The field expert that was consulted saw great potential in this type of application and suggested that the app could be targeting user groups with conditions such as depression, obesity, and chronic diseases. Other research is also suggesting that mHealth apps do have a good effect in terms of promoting physical activity [12].

There was enough data collected through the testing period to document that the various features are working according to the user requirements. Due to the social nature of the features it could be expected that more ideas would be generated by new users and thus leading to new development iterations.

Long term effect of the app has yet to be probed in a different setting. One way would be conducting a proper clinical trial with end points that will look deeper into the efficiency and potentials of helping selected user groups. That will require resources and experts of other fields to work together.

Acknowledgements

The authors acknowledge the expert and end user evaluators and the focus group for their contribution to the research presented in this paper. A special thanks goes to the clinical psychologists who provided her insight into the therapeutic aspects of mHealth in treatment.

References

- [1] WHO | Obesity. (2016). Who.int. Retrieved 18 December 2016, from http://www.who.int/gho/ncd/risk_factors/obesity_text/en/
- [2] Biswas, A., Oh, P., Faulkner, G., Bajaj, R., Silver, M., Mitchell, M. and Alter, D. (2015). Sedentary Time and Its Association With Risk for Disease Incidence, Mortality, and Hospitalization in Adults. *Annals of Internal Medicine*, 162(2), p.123.
- [3] Hevner, A., Salvatore T. M., Jinsoo, P., & Sudha, R. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- [4] Brooke, J. (1996). SUS - A quick and dirty usability scale. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=qjAGPUcAAAAJ&citation_for_view=qjAGPUcAAAAJ:u5HHmVD_uO8C, pp.2-7.
- [5] Jakob Nielsen.Nngroup.com. (2016). 10 Heuristics for User Interface Design: Article by Jakob Nielsen. [online] Available at: <http://www.nngroup.com/articles/ten-usability-heuristics/>
- [6] Knol, A., Slotte, P., van der Sluijs, J. and Lebrete, E. (2010). The use of expert elicitation in environmental health impact assessment: a seven step procedure. *Environmental Health*, 9(1).
- [7] Bryman, A. (2012). *Social research methods*. 4th ed. New York: Oxford University Press, pp.502-504.
- [8] Carpenter, R. (2016). Focus Groups in a Digital Marketing Age. [online] Digital Current. Available at: <http://www.digitalcurrent.com/digital-marketing/focus-groups-in-digital-age/>
- [9] Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction Design: Beyond Human – Computer Interaction* (Vol. 6). Chichester: John Wiley & Sons Ltd. Retrieved from https://www.google.no/books?hl=no&lr=&id=b-v_6BeCwwQC&pgis=1
- [10] Hevner, A., Salvatore T. M., Jinsoo, P., & Sudha, R. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- [11] Fox, K. (1999). The influence of physical activity on mental well-being. *Public Health Nutrition*, 2(3a).
- [12] Fanning, J., Mullen, S. and McAuley, E. (2012). Increasing Physical Activity With Mobile Devices: A Meta-Analysis. *Journal of Medical Internet Research*, 14(6), p.e161.

Address for correspondence

Trond Kjetil Tufte
Email: trondtufte89@gmail.com
Department of Information Science and Media Studies,
University of Bergen,
Fosswinkelsgate 6, 5007 Bergen, Norway.

A Case-Based Study with Radiologists Performing Diagnosis Tasks in Virtual Reality

José Eduardo Venson^{a,b}, Jean Carlo Albiero Berni^b, Carlos Edmilson da Silva Maia^b,
Ana Maria Marques da Silva^c, Marcos Cordeiro d'Ornellas^d, Anderson Maciel^a

^a Instituto de Informática, Universidade Federal do Rio Grande do Sul – Brasil,

^b Animati - Computação Aplicada em Saúde – Brasil

^c Núcleo de Pesquisa em Imagens Médicas, PUCRS – Brasil

^d Departamento de Computação Aplicada, Universidade Federal de Santa Maria – Brasil

Abstract

In radiology diagnosis, medical images are most often visualized slice by slice. At the same time, the visualization based on 3D volumetric rendering of the data is considered useful and has increased its field of application. In this work, we present a case-based study with 16 medical specialists to assess the diagnostic effectiveness of a Virtual Reality interface in fracture identification over 3D volumetric reconstructions. We developed a VR volume viewer compatible with both the Oculus Rift and handheld-based head mounted displays (HMDs). We then performed user experiments to validate the approach in a diagnosis environment. In addition, we assessed the subjects' perception of the 3D reconstruction quality, ease of interaction and ergonomics, and also the users opinion on how VR applications can be useful in healthcare. Among other results, we have found a high level of effectiveness of the VR interface in identifying superficial fractures on head CTs.

Keywords:

Diagnostic Imaging; Image Processing, Computer-Assisted; User-Computer Interface

Introduction

In healthcare, Virtual Reality (VR) has been applied mostly on 3D simulations for training or planning surgeries. Virtual simulators have been especially successful in training minimally invasive and robotic procedures [1]. In diagnosis, however, while 3D image acquisition is ubiquitous (e.g. CT and MRI), the outgoing images are still most often visualized slice by slice or printed out for posterior analysis. One reason is that 2D slices show both internal and external structures in one image. Spatial information, however, is lost. While volume visualization is capable of solving this problem, it faces the difficulty that more internally located structures may be occluded by structures near the surface [2]. In some specific cases only (e.g. planning surgeries and complex fractures), a visualization based on 3D volumetric rendering of the data is considered useful and is currently applied. Even in such cases, the volume is exhibited as an interactive projection on conventional 2D screens and manipulated with mouse and keyboard. On the other hand, the recent popularity of 3D TV and theatres motivated the research on stereoscopic displays for spatial tasks in medical applications [3]. Similarly, the widespread of off-the-shelf VR devices, e.g. Oculus Rift, has attracted the interest of healthcare professionals to a broader range of VR medical application.

The advancement of this area requires a research effort to establish which cases really benefit from a VR interface. Besides, a thorough experimentation will be required to quantify gains for the medical workflow, the patients' health and the healthcare system as a whole.

One key element of using VR in any application is that it renders a comprehensive and intuitive visual representation of the data even for the non-specialist. This opens the possibility to provide exam data to referring physicians that can be used for detailed surgery planning and communication with the patients during medical appointments. Another advantage of VR is that immersion in a virtual environment provides a theoretically unlimited field of view and volume space for the radiologist to organize both 2D and 3D (eventually 4D) data representations, maximizing the compromise between focus and context, increasing the efficiency of the analysis [4]. This is crucial as the profusion of data to be analyzed by the radiologist grows fast as new acquisition modalities and diagnostic techniques evolve.

In a preliminary study with 10 radiologists, we investigated the importance of 3D volumetric reconstructions in the medical workflow. All participants marked the diagnostic option (100%), but only for specific cases such as to search fractures. Other areas have also been mentioned, including surgery planning (90%), educational presentations/classes (70%), operating room, patient appointments and discussion with colleagues (60%). One of the possible innovations in the workflow is the generation of a 3D printed model of the patient specific-anatomy and pathology [5]. However, it is still expensive and limited in terms of modeling soft tissues and model modification. VR is a more flexible and inexpensive alternative.

Given this context, we report, in this paper, a user study to assess VR usage in the diagnostic procedure of fracture identification. Our premise is that VR technology allows for accurate diagnosis with high efficiency. Moreover, we report an exploratory study on the potential of immersive VR for other 3D image-based medical applications, such as virtual endoscopy, surgery planning and patient appointments. The development is focused on two key elements: the first premise is based on the communication between radiologists and referring physicians (e.g. surgeons, cardiologists and orthopedists). As shown in a related work [6], 80 to 90% of referrers agree that the discussion between these professionals improved their understanding of the radiology report, affected patient management, and enhanced radiologists' role. However, as a consequence of an ever-increasing use of teleradiology, the radiologists began working remotely (outside of the clinics and hospitals). Thus, their contact with physicians and patients has substantially reduced. The second premise is that in most cases

the physicians have only the report and raw images, which are poorly comprehensive. Such doctors have knowledge of the patient's clinical case but are dissociated from the imaging result.

Related Work

Ricciardi et al. [7] discussed a medical viewer for 3D environments that can be usable on either desktop, head-mounted display or CAVE. This system allows for inspection of CT and MRI sequences superimposed to the 3D volume made from those images. The software is able to simplify the understanding of complex datasets increasing the visualization realism of anatomical structures by enabling the user's depth perception of the models. No clinical application or evaluation is reported. Similarly, Hänel et al. [8] explore a combination of 2D and 3D images to provide a better understanding of structural changes in the brain of a person with corticobasal syndrome. This system allows for the display on conventional monitors and immersive environments with stereoscopic visualization to improve depth perception. The results show a significant improvement in the spatial localization of brain structures affected with this syndrome.

In diagnosis, King et al. [9] present an immersive virtual reality environment for radiologist work. The study explores a larger screen area provided by VR in comparison with conventional monitors to optimize the volume of images analyzed simultaneously. An application with multiple 2D- only image views was developed on the Unity platform. It is possible to interact with the application through an HMD and to adjust images windowing with a game controller. The system was used for CT visualization of a patient with a lung nodule and multiple-sclerosis lesion evolution of MRI dataset. Validation experiments for the usage of the system in differential diagnosis and remote collaboration were presented. The Oculus Rift was explored in virtual colonoscopy (VC) procedure [10]. A preprocessed mesh is loaded into a virtual environment developed on Unity. Then, a VR camera is assigned to travel both outside and inside the colon, aided by joystick controls. Two radiologists and a gastroenterologist experienced the application, without clinical purposes. The results showed the potential of this technology to improve diagnosis, but emphasize the need of future deployments to provide maximum performance. Likewise, Mirhosseini et al. [11] studied the benefits of 3D immersion for virtual colonoscopy using a CAVE. The authors highlight the benefits of 3D interaction techniques to improve cancer screening in VC.

In these previous studies, the authors present their efforts to establish new techniques of 3D immersive visualization and interaction in medical context. Nevertheless, the VR technology is not widely used in routine medical procedures, such as diagnostic image analysis. Our approach evaluates the diagnostic effectiveness and quality of 3D volumetric reconstructions made for current VR devices.

Methods

In order to measure the capabilities of VR in the diagnostic procedure of fracture identification, we performed a between-subject task-analysis using two CT-scan datasets. Each subject, wearing a head-mounted display and using a couple of interaction tools (windowing and zoom), tried to find fractures in one of the studies. We monitored four dependent variables: the diagnostic effectiveness, 3D images quality, ease of interaction and cyber sickness. This study was approved by the Institutional Review Board of our institution (nbr. 1.782.728).

System Design

We have developed an engine able to apply multiplanar and volumetric reconstructions in tomographic imaging (CT and MRI scans). The system has been designed using the Unity platform and parallel GPU processing.

The visualization engine runs on either a PC with the Oculus Rift or on a smartphone adapted as an HMD using Google Cardboard or other off-the-shelf or 3D printed mobile HMDs. Both platforms allow rotations, translations and windowing (i.e., to select the density corresponding to bone or soft tissue) in the reconstructed volume. A simple switch action, such as a click on a joystick or a trigger on the HMD, allows navigating through the visualization options. Fig. 1 (right) shows users interacting with a joystick to control windowing and zoom. Viewpoint is controlled with natural head rotations to look for a target position in the 3D volume.

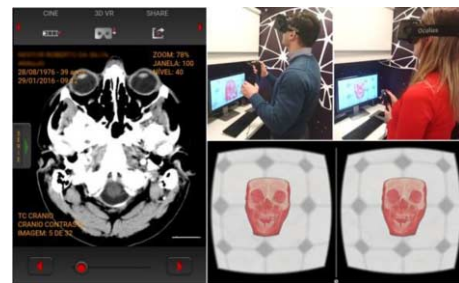


Figure 1 - Integration of the VR interface with PACS. A 3D VR button is shown at the top of the left image. It triggers the 3D VR engine. On the right, we show the users interacting with a 3D dataset through our VR interface¹.

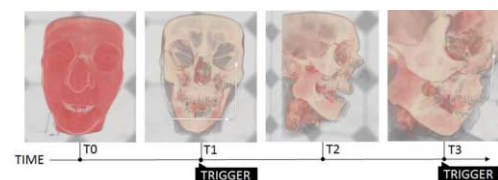


Figure 2 - A typical timeline of user actions with the interface. T_0 shows a centered initial position and windowing in tissue range. At T_1 , the windowing tool is activated through a click and the parameters are set to highlight bones. T_2 shows the volume rotated due to a head movement. At T_3 , the zoom tool is activated to provide details on a specific region.

Fig. 2 shows an example sequence of actions applied with our interface. T_0 is the initial state in which the 3D volume is displayed in the center of the screen with standard windowing and viewpoint. At T_1 a joystick button is pressed to activate the windowing adjustment tool, displaying the two bars with window width and center values. They are changed moving the analog stick in four directions. The result of this change is observed in the second image, highlighting the density range of bone rather than soft tissue. The picture in T_2 represents the user's head movement, which causes a rotation around the object. This natural head movement can be performed until a position of interest is reached. Finally, on time T_3 , a switch action is triggered again, this time enabling the zoom tool. This tool changes the distance to the object using the analog button movement, to focus on a bone fracture for instance.

¹ Please visit <https://goo.gl/uSN47M> to watch a short video demonstration of the interface.

Study Selection and Analysis

We selected two specific exams for our analysis. The exams have been searched in the partner clinic's database with the strings: *comminuted fractures*, *bone fracture* and *polytrauma*. Head CT studies appeared as a convenient choice as they are common in routine examination of trauma situations. Moreover, they are suitable for analysis by any radiologist, regardless of specialty. The two head CT studies with the highest number of fractures were selected and anonymized. The first one (a) with 191 slices, thickness 1 mm and bone filter applied in the acquisition. The second dataset (b) has 231 slices and the same thickness of the first.

We conducted a preliminary interpretation to identify existing fractures in each data-set. Three neuroradiologists with 5 to 10 years of experience performed this analysis. Clinical findings were obtained by viewing images in the conventional way (i.e. desktop computer and 2D monitor), using slices in the axial, sagittal and coronal planes, in conjunction with the 3D volume. Seven fractures have been identified in the dataset *a*, and six in dataset *b* by the committee of 3 experts. Doctors emphasized that some of these fractures are internal and small, hardly interpreted only viewing the 3D volume as they may be occluded by structures near the surface. After the studies interpretation, we prepared two lists with the clinical findings of each dataset. In each list, we included two additional fractures that do not exist to provide false alternatives. In some of them, we simply reversed the laterality of fracture. In other cases, false alternatives pointed to non-existing fractures in locations near existing ones.

Experimental Setup

Fifteen radiologists and one medical physicist participated voluntarily in the study (thirteen males and three females). Their times of experience ranged from 1 to 5 years (56.3%), 5 to 10 (25%) and over 10 years (18.8%). The subjects have subspecialties in radiology. The 33.3% are members of the neuroradiology team, followed by 26.7% of specialists in abdominal radiology, 26.7% in musculoskeletal, 6.7% in thoracic and 6.7% in other subspecialties. This study was the first experience with virtual reality for 81.3% of participants. The main user task in our experiment is to find fractures in a reconstructed 3D volume in an immersive environment by applying transformations (such as zoom and rotation) and windowing adjustment. No additional clinical information or viewing in 2D planes were available. Furthermore, no cutting planes were allowed in this version of the interface to prevent users to simulate slice-by-slice visualizations.

Initially, each participant signed a legal term of consent informing about the compliance with ethical precepts and the details of the experimental protocol. Before starting the experiment, all the subjects also filled out a characterization questionnaire. The following step was an interface learning session in which a video demonstrating the use of the application and methods of interaction with the HMD and joystick was presented. No preliminary training was allowed. For this experiment, we chose to use the Oculus Rift DK2 attached to a desktop computer. This choice has been motivated by the wider availability of the device, which can be useful for comparison by other researchers, and due to the device specifications, e.g. higher resolution and performance when compared with Google Cardboard interface. The resolution displayed to each eye by the Oculus Rift DK2 is 960 x 1080 pixels, refresh rate is 75Hz, and field of view is 100°.

After the learning session, the user performed the task of finding fractures in one dataset for 2 to 5 minutes according to their preference. Eight users viewed dataset *a* and the other 8 analyzed dataset *b*. Finally, subjects filled out a post-questionnaire with questions about the list of possible fractures

found, ease of interaction and comfort during the task. For the comfort analysis, we adapted the Simulator Sickness Questionnaire (SSQ) [12], using the questions most related to our case. The list of possible fractures has been defined in the preliminary interpretation with true and false alternatives. Participants can select as many alternatives as they wish.

Results

The results are presented here. Authors may choose a combination of text, tables, figures, and graphs to convey the results of their work to the reader. There are no set limitations on the number of tables, figures, and graphs that may be used in papers, posters, and proposals. Large figures and tables may span two columns. Please number tables and figures and reference them appropriately in the text.

We report results on the users' performance in identifying fractures, the subjective quality of the reconstructions and subjective comfort/discomfort.

Diagnostic effectiveness

Figs. 3 and 4 show the number of responses on each of the possible observations on the datasets *a* and *b* respectively. The options with marker * represent the false alternatives and the maximum number of answers by alternative is eight.

The first exam has 7 correct answers and 2 false, as illustrated in Fig. 4. Regarding the correct options, 7 radiologists (87.5%) correctly interpreted the lateral right orbit wall and right orbit floor. The right zygomatic arch has been found by 5 radiologists (62.5%), the nasal bones 4 (50%) and maxillary sinus walls 3 (37.5%). The left orbit floor and Le Fort type I had a single vote each. *Le Fort* indicates fractures of the midface, which collectively involve separation of all or a portion of the midface from the skull base. It is a more specific classification of neuroradiology and not commonly used by other areas. Notice that for the dataset *a*, all true fractures were identified by at least one radiologist. The answers to the two false alternatives are also shown in the chart of Fig. 3. No subject marked the left zygomatic arch (100% hit), because this fracture was on the right side and VR viewing had no impact on the correctness of laterality. Mandible (jawbone) ramus had one wrong vote made by the only participant who is a medical physicist and is not expected to succeed in identifying fractures.

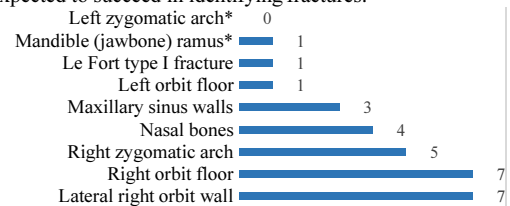


Figure 3 - Diagnostic effectiveness for dataset *a*. Possible fractures are in the y-axis. The x-axis shows the number of users that have marked each possible option. The * indicates that the respective bar represents a non-existent fracture.

Fig. 4 shows the answers on the second exam (b) with six correct and two false choices. All radiologists found the depressed right anterior maxillary sinus wall (100% hit). Then, the right zygomatic arch and depressed right anterior frontal sinus wall received 7 votes each (87.5% hit). The posterior maxillary sinus wall, which is an internal structure, scored only two votes (25%) and other 2 correct fractures in internal structures have no vote. This was expected, as the ethmoidal cells, posterior frontal sinus wall and posterior maxillary sinus wall are internal head structures, hardly seen without making

cuts in the 3D volume or viewing on 2D slices. In the preliminary interpretation, the radiologists explicitly pointed to this possibility, and these results demonstrate that.

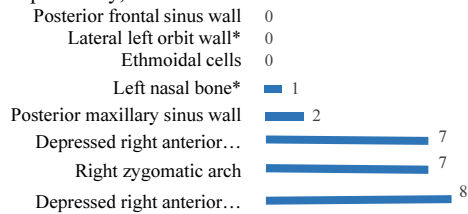


Figure 4 - Diagnostic effectiveness for dataset b. Possible fractures are in the y-axis. The x-axis shows the number of users that have marked each possible option. The * indicates that the respective bar represents a non-existent fracture.

Similarly, to the first exam, the dataset *b* scored only one vote in one of the false alternatives (left nasal bone), and none in the other. The exam presents facial polytrauma affecting nearby structures. This seems to have induced a participant to precipitate (rather a memory effect than perception). The lateral left orbit wall had no marking, as expected.

As we hypothesized, these results showed a high accuracy rate identifying superficial fractures in both studies. Most radiologists, regardless of experience and specialty, found the outer skull fractures, for instance, in the zygomatic arch, anterior maxillary sinus wall and lateral orbit wall. Thus, we emphasize the high diagnostic effectiveness obtained with the VR interface, even though it was the first contact of the subjects with an immersive VR system. Results have also shown a limitation for identification of the internal fractures, which was expected as they are occluded by surface elements.

Usability and Quality

We also analyzed two variables related to our interface: ease of interaction and quality of 3D reconstructions. The subjects graded the easiness of interacting using our interface compared to other interfaces for 3D data they are familiar with, e.g. their conventional workstations where mouse and keyboard are used to apply transformations in volumetric images. They also graded the quality of the 3D reconstructions in comparison with the conventional volumetric images, which most often are preprocessed on the acquisition equipment and then sent as 2D images for the radiologists. The responses were collected in a five points Likert Scale questionnaire, from 1 (very low) to 5 (very high).

Fig. 5 shows the percentage of selections in each level for quality and ease of use (usability) for both datasets. Concerning usability, the two studies had no negative selections. The general average was 4.0625 points (max = 5, min = 2, stdev = 1.44). This assessment reflects the simplicity of our interface to focus on a region (zoom + head rotations) and to set parameters of interest (window).

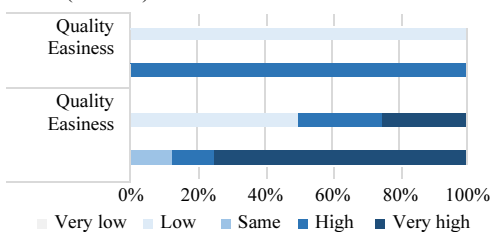


Figure 5 - The bars represent the percentage of user selections for quality and easiness grouped by dataset. The answers range from very low (light blue) to very high (dark blue).

Concerning the quality of the reconstructed image volume, we rely on the subjective analysis of radiologists and medical physicists, who are very demanding in terms of image attributes. In both studies their answers present a general average of 2.85 (max = 5, min = 2, stdev = 1.1474). The visualization of dataset *a* (avg = 3.125) has been perceived as of better quality than the dataset *b* (avg = 2.65), even with a smaller number of slices. This probably occurred due to the bone filter used in image acquisition, highlighting the surface structures and consequently generating a sharper image.

Notice that the image quality is impaired by the relatively low resolution of the Oculus Rift DK2. Besides, the algorithms have been optimized to run on mobile devices while most 3D images currently available to radiologists are processed on dedicated workstations. Despite these limitations, they were able to diagnose the surface fractures of the volume, as previously shown.

Exploratory Analysis and Discomforts

Another contribution of our research is the user's opinion about where VR applications can be useful in medicine. According to 75% of radiologists, a VR interface such as ours has benefits in virtual endoscopies, like colonoscopy, where the conventional procedure is the use of mouse and keyboard to navigate in a 3D volume displayed on a 2D monitor.

The 68.8% of the participants think this interface would be useful for communication with referring physicians. The doctors explained that medical assistants have difficulty understanding the 2D slices. They are unable to make an appropriate mental spatial reconstruction. Recent work [6] has shown that promoting communication between the radiologist and the referring physician improved the understanding of the images and the exams report, consequently affecting patient healthcare. Our results highlight the potential of VR to analyze 3D images with intuitive visual representations that instigate discussion among the medical staff. This is particularly important in planning surgery and analyzing complex fractures, as highlighted by 75% of subjects.

We then explored further, asking the participants what tools they would like to see integrated into the interface to help them in diagnosis. Most of them (62.5%) indicated that oblique cuts in the volume (in arbitrary planes) would be useful. This is reported to be crucial to visualize internal structures. The use of transparency in the mapping of windowing parameters to the volume rendering helps seeing internal structures. However, it is known that humans do not deal well with several transparent layers [13]. Besides, 10 physicians (62.5%) would like to be able to navigate in 2D planes with reference lines/planes displayed in the volume. VR provides a 360° wide visualization space that enables volume and planar views to be placed side by side.

Finally, eventual discomforts with the interface were measured with an adapted Simulator Sickness Questionnaire. All participants performed the task standing upright and without prescription glasses. Fig. 6 shows the distribution of the discomforts reported by the subjects. Overall, a very low level of discomfort was reported. Blurred vision was the main issue, affecting nearly 25% of the users. Similarly, 3 participants reported eyestrain (2 moderate and 1 severe). These are linked to 5 users that reported some degree of myopia. In a day-by-day workflow with a VR interface, they should be allowed to wear their prescription glasses or adjust the HMD optics individually.

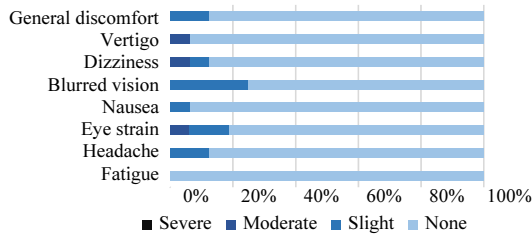


Figure 6 - Each bar shows the intensity of each evaluated discomfort during the execution of the fracture identification tasks. The overall discomfort is surprisingly low.

Conclusion

In this paper, we presented a user study with medical specialists to assess diagnostic effectiveness of VR usage in fracture identification. We performed experiments to validate the proposed approaches with sixteen expert professionals in image diagnostic procedures. Subjects were challenged to identify fractures in head CT exams in a virtual environment.

The results have shown high effectiveness in identifying superficial fractures for two different volume exams. However, we found that viewing only the volume surface is not enough for the complete diagnosis, as deeply located internal structures are hard to visualize on the reconstructed volume. These results support our premise of VR being suitable to provide a more intuitive interface for the whole chain of the medical care. One remarkable observation is that it may encourage communication between referring physicians and radiologists as it increases the exam comprehension by the non-specialist. Increased communication is often correlated with better decision-making [6].

Future work should focus on the assessment of long term use of VR goggles for image-based diagnosis.

Acknowledgements

The study is part of the project coordinated by the company Animati - Computing for Healthcare and funded by TECNOVA-RS FAPERGS/FINEP (1451-2551/14-6) program. Authors are also supported by CNPq grant 305071/2012-2, and FAPERGS project 2283-2551/14-8.

References

- [1] van Beurden, M. H., IJsselstein, W. A., & Juola, J. F. (2012). Effectiveness of stereoscopic displays in medicine: a review. *3D Research*, 3(1), 1-13.
- [2] Hänel, C., Pieperhoff, P., Hentschel, B., Amunts, K., & Kuhlen, T. (2014). Interactive 3D visualization of structural changes in the brain of a person with corticobasal syndrome. *Frontiers in neuroinformatics*, 8.
- [3] Escobar, M. M., Junke, B., Holub, J., Hisley, K., Eliot, D., & Winer, E. (2015). Evaluation of monoscopic and stereoscopic displays for visual-spatial tasks in medical contexts. *Computers in biology and medicine*, 61, 138-143.
- [4] Tyndiuk, F., Lespinet-Najib, V., Thomas, G., & Schlick, C. (2004, November). Impact of large displays on virtual reality task performance. In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa* (pp. 61-65). ACM.

- [5] Rengier, F., Mehndiratta, A., von Tengg-Kobligh, H., Zechmann, C. M., Unterhinninghofen, R., Kauczor, H. U., & Giesel, F. L. (2010). 3D printing based on imaging data: review of medical applications. *International journal of computer assisted radiology and surgery*, 5(4), 335-341.
- [6] Rosenkrantz, A. B., Sherwin, J., Prithiani, C. P., Ostrow, D., & Recht, M. P. (2016). Technology-Assisted Virtual Consultation for Medical Imaging. *Journal of the American College of Radiology*.
- [7] Ricciardi, F., Pastorelli, E., De Paolis, L. T., & Herrmann, H. (2015, August). Scalable medical viewer for virtual reality environments. In *International Conference on Augmented and Virtual Reality* (pp. 233-243). Springer International Publishing.
- [8] Hänel, C., Pieperhoff, P., Hentschel, B., Amunts, K., & Kuhlen, T. (2014). Interactive 3D visualization of structural changes in the brain of a person with corticobasal syndrome. *Frontiers in neuroinformatics*, 8.
- [9] King, F., Jayender, J., Bhagavatula, S. K., Shyn, P. B., Pieper, S., Kapur, T., ... & Fichtinger, G. (2016). An immersive virtual reality environment for diagnostic imaging. *Journal of Medical Robotics Research*, 1(01), 1640003.
- [10] Randall, D., Metherall, P., Bardhan, K. D., Spencer, P., Gillott, R., de Noronha, R., & Fenner, J. W. (2015). The Oculus Rift virtual colonoscopy: introducing a new technology and initial impressions. *Journal of Biomedical Graphics and Computing*, 6(1), p34.
- [11] Mirhosseini, K., Sun, Q., Gurijala, K. C., Laha, B., & Kaufman, A. E. (2014, November). Benefits of 3D immersion for virtual colonoscopy. In *3DVis (3DVis), 2014 IEEE VIS International Workshop on* (pp. 75-79). IEEE.
- [12] Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3), 203-220.
- [13] Kawabe, T., Maruya, K., & Nishida, S. Y. (2015). Perceptual transparency from image deformation. *Proceedings of the National Academy of Sciences*, 112(33), E4620-E4627.

Address for correspondence

José Eduardo Venson, jeveson@inf.ufrgs.br, +55 51 999087357

Preferred Features of E-Mental Health Programs for Prevention of Major Depression in Male Workers: Results from a Canadian National Survey

JianLi Wang^a and Kendall Ho^b

^a Institute of Mental Health Research, University of Ottawa, Ottawa, Ontario, Canada,

^b Digital Emergency Medicine, University of British Columbia, Vancouver, British Columbia, Canada

Abstract

The purpose of this study was to estimate and compare the proportions of preferred design features and likely use of e-mental health programs and understand potential barriers to the use of e-mental health programs in working men who were at high risk of a major depressive episode. A cross-sectional survey in 10 provinces in Canada was conducted between March and December 2015 (n = 511). Of the 17 different features assessed, the top three features that were most likely to be used by high-risk men were: “information about improving sleep hygiene” (61.3%), “practice and exercise to help reduce symptoms of stress and depression” (59.5%) and “having access to quality information and resources about work stress issues” (57.8%). E-mental health programs may be a promising strategy for prevention of depression in working men. Development of e-mental health programs should consider men’s preferences and perceived barriers to enhance the acceptability of this approach.

Keywords:

Mental Health Services; Program Development; Depressive Disorder, Major

Introduction

Major depression is a prevalent mental disorder. In Canada, the annual prevalence of major depressive disorder was 3.9% based on the DSM-IV criteria [1]; in the United States, the 12-month prevalence was 6.6% [2]. Major depression is also disabling and significantly affects workers’ health and productivity. Depression alone accounts for 2.5% of the global burden of disease and is among the largest single causes of disability worldwide (8.2% of all years lived with disability globally) [3]. US workers with depression cost an estimated US \$44.01 billion per year in lost productivity [4]. Epidemiological studies have consistently found that women are more likely to have major depression than men. However, men are not immune to depression. Canadian national data showed that annually, 2.8% of adult men have a major depressive disorder [1]. One of the severe consequences of having a major depressive episode (MDE) is suicide. Canadian national data showed that 75%-80% of all suicides were male [5, 6].

Given the considerable impact on health, productivity and lives, there is a pressing need for innovation in the prevention of major depression in male workers. However, this is a challenging endeavor. In the workplace, risk factors for having MDE differ for men and women [7-9]. For instance, job strain,

family to work conflict and job insecurity appear to be more prominent MDE risk factors in men than in women [7, 8]. Compounding men’s risk, men are less likely than women to seek help and to disclose depressive symptoms and often delay seeking help until symptoms become severe. Men tend to be socialized to be emotionally stoic and exemplify traditional masculine characteristics such as independence, self-reliance, and dominance [10]. Men are concerned over the perceived negative judgments from family, friends, and co-workers if they access treatment for depression. These gender-specific experiences together with a limited knowledge base about effective interventions call for innovative solutions tailored for men. One of the burgeoning solutions that have attracted considerable interests is e-mental health.

E-mental health is “the use of information and communication technologies to support and improve mental health, including the use of online resources, social media and smartphone applications” [11]. E-mental health self-help services enable users to learn more about their mental health conditions and empower users to strengthen their self-management and improve their health. The majority of e-mental health treatment programs were developed based on the models of cognitive behavioral therapy (CBT) or interpersonal therapy (IPT) that demonstrate evidence of effectiveness for depression and anxiety [12]. A review by Christensen and Petrie showed that by 2013 there had been 62 web-based mental health interventions and 11 mobile applications [13]. Lal and Adair found 91 peer-reviewed publications on the application of e-mental health interventions between 2000 and 2010. Thus far, randomized controlled trials of e-mental health interventions on depression have been conducted with samples of individuals who have clinical depression or are above a threshold of a depression rating scale [14]. Consistent with public health mandates to reduce the burden of depression, it is important that e-mental health not only addresses needs of those with depression or who are above a depression rating threshold, but also advances capacity for secondary prevention [15], i.e., identifying high-risk individuals and intervening to prevent early symptoms from progressing into an MDE. Additionally, existing e-mental health programs have not been designed and evaluated using a gender lens. Given the gender norms, the extent to which e-mental health programs are accepted by men is unknown. The effectiveness of the program and its acceptability to users constitute the foundation for scalable and sustainable program implementation. Therefore, as part of the BroMatters study (www.bromatters.ca), we conducted a cross-sectional survey in working men, some of whom were at high risk of MDE, to understand their preference of design features of e-mental health programs.

The objectives of this analysis were, amongst male workers who were at variable risk levels of MDE, to 1) estimate and compare the proportions of Internet use for medical information, preferred design features and likely use of e-mental health programs, 2) examine factors associated with the likely use of e-mental health programs, and 3) understand potential barriers to the use of e-mental health programs in working men who were at high risk of MDE.

Methods

A cross-sectional survey was conducted between March and December 2015. The target population of the survey included Canadian working men who: 1) were aged 18+, 2) did not have an MDE in the past 12 months, 3) were at high risk of MDE at the time of interview (a low-risk sample was also obtained for comparison), 4) were working at the time of the survey, and 5) have no language barriers to either English or French. Because of the vast geographic area of Canada, participants were recruited using random digit dialing method by Bureau of Professional Interviewers (BIP) located in Montreal Canada. The BIP has the access to household telephone numbers across the country and to a validated cell phone number database, and its interviewers can conduct interviews in both English and French. The study was approved by the Conjoint Health Research Ethics Review Board of University of Calgary.

Once a household was reached, the household contact was asked to retrieve, or provide contact information (e.g. a first name) of the household residents who are men and are currently working. If there was more than one potentially eligible individual in the same household, one was randomly selected. Once the prospective participant was fully informed about the objectives and procedures of the study, oral consent was obtained to proceed with the interview. Participants were first administered a risk calculator for MDE to estimate their probability of having MDE in the future. The definition of high-risk is described below. The number of high risk participants in each age group was proportional to the age distribution of Canadian male working population in 2014, provided by Statistics Canada.

From March to December 2015, 49,500 calls were made. A majority of the calls (47,648, 96.2%) were not valid (not in service, fax/modem, answering machine, language barriers, ineligibility, duplications, refusal before eligibility was assessed). Among 1852 eligible participants, 596 (32.1%) refused to participate after verifying eligibility; 842 provided completed data (45.4%); the remaining included incomplete interviews and scheduled call-backs not in the study period (22.5%). After removing one duplication, 841 participants were included in the analysis, including 511 men who were at high risk of having major depression and 330 who were at low risk of having major depression.

Measurements

A *multivariable risk prediction algorithm for major depression in men* was administered to estimate the risk (probability) of having MDE in the next 4 years for each participant [16]. This risk prediction model was designed to be used in individuals who do not have an MDE. Based on participant's exposure to a key set of risk factors (predictors) in the model, the algorithm can generate the absolute risk/probability of having MDE in the next 4 years, analogous to the Framingham risk prediction algorithm for coronary heart disease [17, 18]. The risk prediction algorithm for MDE

in men was developed and validated using data from 4737 Canadian men who were aged 18+ years and who did not have MDE in the past 12 months [16]. The risk prediction algorithm contains 15 predictors including age, personal and family history of MDE, childhood trauma, ongoing stress and life events, and anti-depressant or sleeping pill use in the past month. The predictive power of the risk prediction algorithm was measured by C statistics ($C = 0.7953$) [16] which is equivalent with area under the curve when the outcome is binary. The model had excellent calibration with data as indicated by the Hosmer-Lemeshow test and visual comparison between the predicted and observed risks by decile risk groups [16]. In our study, 6.51+% were defined as high risk for men, which represents the top two decile risk groups in the Canadian male population. The predicted risk that was lower than 6.51% was defined as low risk.

Internet use was assessed using questions from the 2012 Canadian Internet Use Survey conducted by Statistics Canada.

Preferred design features of e-mental health program questions were developed by the BroMatters team members. Participants were asked "We want to hear your opinion about e-mental health programs for dealing with work and stress issues. E-health is defined as For the following features, please indicate how likely it is that you would use them." Seventeen questions about design features were asked. For each question, participants answered on a 5-point Likert scale ranging from: very likely to very unlikely. Open-ended questions were asked about any other features they may want in an e-mental health program, whether the participant and his male co-workers may use an e-mental health program to deal with work stress, and what make it difficult to use an e-mental health program.

Ineligible participants, administering the questions and instruments took an average of 22 minutes to complete. Participants who completed the survey received a CAN \$20 gift card as a token of appreciation.

Statistical analysis

The background characteristics and proportions of likely use of design features were estimated and compared in men who were at high or low-risk levels of having MDE, using Chi square test. In men who were at high-risk of having MDE, the percentages were also estimated and compared by age groups and by language used in the interview (English vs. French), using Chi-square test. All analyses were conducted using the statistical program STATA 14.0 [16]. Tests were considered statistically significant when p was < 0.05 . With this level of probability and a sample size of 841, the study had a level of statistical power of 0.89 to detect a small effect size (Cohen's d) of 0.20.

Results

A majority of the participants reported use of Internet for personal reasons in the past 12 months, with the proportion (95.7%) in the high risk men slightly higher than that in the low risk men (92.4%). The two groups did not differ in Internet use for searching medical information and in perceived usefulness of the Internet information in making decisions about health. However, high risk men (83.4%) were more likely to have reported that it was important to access health resources on the Internet than low risk men (75.0%).

Participants rated their level of interest in possible use of 17 different features that can be incorporated into the design of e-

mental health programs. We ranked the preferred design features of e-mental health program in descending order. The top three features that were identified by high-risk men as things they would likely to use were: *“information about improving sleep hygiene”*, *“practice and exercise to help reduce symptoms of stress and depression”* and *“having access to quality information and resources about work stress issues”*. The proportions of individuals endorsing the selected design features in the high-risk group were significantly higher than those in the low-risk group, except for the feature of *“information about improving sleep hygiene”*.

We estimated and compared the proportions of preferred design features by age groups and languages used in the interview in men who were at high risk of MDE. The data showed that, compared to older participants (aged 65+ years old), younger participants preferred access to a program through smartphone or mobile applications and that the information be delivered in game format. Middle aged men (aged 30 to 64 years old) had a preference for receiving printed materials. The preferences for other design features did not vary by age.

English-speaking participants were more likely to use *“practice & exercise to reduce stress”*, *“access a program via smartphone or an app.”*, and *“being able to ask questions and receive answers from mental health professionals”* than French-speaking men; French-speaking men were more likely to use *“being able to chart and track your mood”* than English-speaking participants.

A majority of participants considered our survey questions about preferred design features to be comprehensive and did not have other features to add. For the open-ended questions, some participants suggested that, in addition to the design features encompassed in the survey, other valuable features of e-mental health programs may include: easy to use (e.g., *“online information in a format that's simple to use.”*), confidentiality (e.g., *“Privacy, somehow to ensure privacy”*), credibility (e.g., *“having access to, to reliable information that's important to me”*) and direct link to a professional (e.g., *“like some kind of call in line. Like a hotline where you could access a live expert..... something personal”*). (Note: *italics* denote direct quotes.)

Among the participants, 69.0% reported “yes” or “maybe” to potentially using an e-mental health program to deal with work stress. The percentage was higher in the high-risk group (72.6%), those in the levels of higher education and personal income, younger age groups, and those working in mid and large companies, compared with their counterparts. No differences were found by language, marital status and employment status (employee vs. self-employed). High MDE risk participants who reported that they would not use an e-mental health program for dealing with stress were asked: “what would make it difficult to use an e-health program?” The reported barriers included perceived stigma associated with accessing e-mental health support (e.g., *“...social stigma, comfort of access”*, *“...workplace ignorance and what do they call that where you stereotype ...”*), lack of personal interaction inherent to e-mental health (e.g., *“lack of personal face to face”*, *“... don't see the value of it if you could talk to your family doctor...”*), lack of time (e.g., *“...if it was time consuming...”*), and lack of knowledge (e.g., *“Well the fact that I don't know what an e-health program is, makes it difficult. I'm not sure that (laughs)”*). (Note: *italics* denote direct quotes.)

Discussion

This study is descriptive in nature as the results were used for guiding the development of e-mental health program. One key finding of this study was that 62.7% participants who were at high risk of having MDE had used the Internet for health information in the 12 months prior to the survey. This percent is slightly higher than a similar estimate from the 2012 Canadian Internet Use survey, in which 60.8% men reported use of Internet for medical or health related information [20]. Furthermore, over 75% of high MDE risk men in our sample considered health information on the Internet to be useful in helping them make health decisions and over 72% would use an e-mental health program to deal with work-related stress. Given that men often delay help-seeking for mental health problems because of perceived stigma and gender norms, our results suggest that the privacy inherent to e-mental health programs make e-mental health programs be a promising tool for improving men's mental health.

Acceptability of a tool is vital to the evaluation of its effectiveness and implementation. Therefore, to develop e-mental health programs for men, it is critical to understand their preferred design features. It is interesting to observe, in our survey, that *“information about improving sleep hygiene”* was the top design feature preferred by men, irrespective of their risk status. Individuals who are at high risk of MDE may be occupied by unhelpful thinking and look for strategies to solve the issues they encounter. Thus, it is not surprising that the second top feature they endorsed was *“practice and exercise to help reduce symptoms of stress and depression”* which is consistent with the principles of CBT, e.g., changing unhelpful thinking and behaviors, and problem-solving focused. We anticipated that CBT practices and educational information (*“having access to quality information and resources about work stress issues”*) would be needed by the participants and this was demonstrated in this study. This also is consistent with the fact that most of the existing e-mental health programs (such as MoodGYM [21]) were developed based on the CBT approach [22]. We found that men who were at high risk of having MDE were more likely to have endorsed the design features than men who were at low risk. No age differences were found in preferred design features. English-speaking participants were more likely to use CBT techniques and an app and French-speaking participants were more likely to use mood monitoring tools. These results indicate that e-mental health programs incorporating these preferred features are likely to be used by men who are at high risk of MDE across age and English/French speaking categories.

Comparison with Prior Work

An understanding of barriers to the use of e-mental health programs is also important for the development, evaluation and implementation of the programs. Some features preferred by the participants reflect the concerns they have about e-mental health programs and potential barriers to the use. Based on the quantitative and qualitative data, confidentiality and privacy protection are the prominent concerns for high-risk participants. They were concerned about the consequences if others know that they use the program to deal with stress related issues (e.g., perceived stigma). Other barriers include the extent to which the program is easy to use and navigate, credible (e.g., information is provided by health professionals), relevant to one's personal situation, and interactive (e.g., being able to communicate with a

professional). Additionally, lack of time and knowledge about e-health are potential barriers reported by the participants. Schneider and colleagues investigated users' views of an online CBT program (MoodGYM) in a randomized controlled trial [23]. Wetterlin et al.'s cross-sectional study examined youth expectations for mental health websites [24]. Both studies reported preferences and perceived barriers that are consistent with the results of our survey.

Our study has several limitations. First, the survey data relied on self-report. Therefore, reporting and recall biases are possible. Second, our target population is Canadian working men who were aged 18 and older. Compared to men in the Canadian workforce in 2014, our sample was slightly older. The proportion of our participants aged 18 to 29 was 12.1%, while it was 20.2% in the Canadian workforce. Therefore, the proportions of some design features by age groups could have been over or underestimated due to potential selection bias. Given the increasing use of cell phones in young adults, future studies may investigate strategies for recruiting young adults through cell phones. Finally, our survey collected self-reported qualitative information about barriers to the use of e-mental health programs. The qualitative information should be considered preliminary. More studies are needed to provide definitive answers.

Conclusions

There is a pressing need for developing innovative strategies for prevention of depression in men. This is a challenging endeavor because of the gender norms and social stigma against depression and help-seeking among men. E-mental health holds potential as it can be confidential, easily accessible and economic if it is appropriately designed. However, the design of e-mental health solutions should consider the sex and age differences in terms of preferred features. More studies are needed to examine preferred design features and the barriers to use in different populations so that e-mental health strategies that meet the needs of different age groups and personal background can be developed.

Acknowledgements

This study was supported by a team grant from Movember Foundation.

References

- [1] Patten SB, Williams J, Lavorato DH, Wang JL, McDonald K, Bulloch AG. Descriptive epidemiology of major depressive disorder in Canada in 2012. *Can J Psychiatry* 2015; 60:23-30. PMID:2681717
- [2] Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; 289:3095-3105. PMID:24081188
- [3] Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, Vos T, Whiteford HA. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med* 2013; 10:e1001547. PMID: 24223526
- [4] Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA* 2003; 289:3135-3144. PMID:22556191
- [5] Smith ND, Kawachi I. State-level social capital and suicide mortality in the 50 U.S. states. *Soc Sci Med* 2014; 120:269-277. PMID: 25108693
- [6] Statistics Canada. Suicide rates, an overview, 1950 to 2009. 2014. <http://www.statcan.gc.ca/daily-quotidien/120725/dq120725a-eng.pdf> <http://www.webcitation.org/6gCZzU4W0>
- [7] Wang J, Patten SB, Currie S, Sareen J, Schmitz N. A Population-based Longitudinal Study on Work Environmental Factors and the Risk of Major Depressive Disorder. *Am J Epidemiol* 2012; 176:52-59. PMID:19318611
- [8] Wang J, Schmitz N, Dewa C, Stansfeld S. Changes in perceived job strain and the risk of major depression: results from a population-based longitudinal study. *Am J Epidemiol* 2009; 169:1085-1091. PMID:18079332
- [9] Wang JL, Lesage A, Schmitz N, Drapeau A. The relationship between work stress and mental disorders in men and women: findings from a population-based study. *J Epidemiol Community Health* 2008; 62:42-47. PMID:23790813
- [10] Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. *Am Psychol* 2003; 58:5-14. PMID: 12674814
- [11] NHS Confederation. *E-Mental Health: What's all the Fuss About?* 2014. <http://www.nhsconfed.org/-/media/Confederation/Files/Publications/Documents/E-mental-health.pdf> <http://www.webcitation.org/6gCakiYcz>
- [12] Cuijpers P, Donker T, van SA, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. *Psychol Med* 2010; 40:1943-1957. PMID:20406528
- [13] Christensen H, Petrie K. State of the e-mental health field in Australia: where are we now? *Aust N Z J Psychiatry* 2013; 47:117-120. PMID:23297367
- [14] Lal S, Adair CE. E-mental health: a rapid review of the literature. *Psychiatr Serv* 2014; 65:24-32. PMID:2404268
- [15] Andrews G. Implications for intervention and prevention from the New Zealand and Australian mental health surveys. *Aust N Z J Psychiatry* 2006; 40:827-829. PMID: 16959007
- [16] Wang JL, Manuel D, Williams J, Schmitz N, Gilmour H, Patten S, MacQueen G, Birney A. Development and validation of prediction algorithms for major depressive episode in the general population. *J Affect Disord* 2013; 151:39-45. PMID:25519847
- [17] D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Masaro JM, Kannel WB. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; 117:743-753. PMID:18212285
- [18] Lloyd-Jones DM. Cardiovascular risk prediction: basic concepts, current status, and future directions. *Circulation* 2010; 121:1768-1777. PMID: 20404268
- [19] StataCorp. *Stata Statistical Software: Release 14.0*. 2015. College Station, TX: Stata Corporation. (Computer Program)
- [20] Statistics Canada. Canadian Internet Use Survey. 2013. <http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=130941> <http://www.webcitation.org/6gCZ4Z4aW>
- [21] Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. *J Med Internet Res* 2002; 4:e3. PMID:11956035
- [22] Gratzner D, Khalid-Khan F. Internet-delivered cognitive behavioural therapy in the treatment of psychiatric illness. *CMAJ* 2016; 188:263-272. PMID:12813115
- [23] Schneider J, Sarjami FP, Grime P, Thormicroft G. Acceptability of online self-help to people with depression: users' views of MoodGYM versus informational websites. *J Med Internet Res* 2014; 16:e90. PMID:25262314
- [24] Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. *J Med Internet Res* 2014; 16:e293. PMID:23993280

Address for correspondence

JianLi Wang

Email: jlwang@ucalgary.ca

Development of a Smartphone Application to Monitor Pediatric Patient-Reported Outcomes

Jingting Wang^a, Nengliang(Aaron) Yao^b, Yanyan Liu^a, Zhaohui Geng^a, Yuanyuan Wang^a, Nanpin Shen^c, Xiaoyan Zhang^d, Min Shen^e, Changrong Yuan^a

^a School of Nursing, Second Military Medical University, Shanghai, China,

^b Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

^c Department of Nursing, Shanghai Children's Medical Center, Shanghai, China

^d Department of Hematology and Oncology, Shanghai Children's Medical Center, Shanghai, China

^e Department of Hematology, Children's Hospital of Soochow University, Jiangsu, China

Abstract

This paper is to document the development process of a smartphone app to track patient-reported outcomes (PROs) of children with chronic diseases, and to share the usability study results. A multidisciplinary research team including health services researchers, pediatric nurses, and software engineers worked synergistically in development. Group discussions and several rounds of feedback and modification were conducted. Ten pediatric patients, five parents, and two nurses participated in usability study. Qualitative content analyses were employed in development and usability evaluation. The app collects demographic information and PROs. The web-based administration portal helps to manage demographic information, questionnaires, administrators, and survey-conducting organizations. All participants felt this app was easy to use and the interfaces were friendly to children. Nurses thought administration portal interfaces were simple and data was convenient to download for further analysis. The app and administration portal have potential to promote PROs in assessing quality of life and symptoms in pediatric patients.

Keywords:

Patients, Mobile applications, Software design.

Introduction

The Patient-Reported Outcomes Measurement Information System (PROMIS) Pediatric Cooperative Group has developed pediatric self-report item banks for youth ages 8-17 years across five generic health domains (physical functioning, pain, fatigue, emotional health, and social health) [1, 2]. Our research team has been authorized to translate the PROMIS Pediatric and PROMIS parent proxy report short forms into Simplified Chinese. Pediatric PROMIS contains instruments for 8- to 17-year-olds [1, 2]. PROMIS Parent Proxy Report Scales for Children were developed for parents of children ages 5-17 years who were too young, too cognitively impaired, or too ill to complete a PRO instrument [3-5]. Computer-assisted tools have been developed in recent years to collect PROs. However, computer-based tools are often only used in hospitals or clinics, which makes it hard to collect PROs when patients are at home in the break before the next treatment. Smartphone applications (apps) have the potential to collect PRO data with high efficiency and effectiveness. Our research team developed a smartphone app to track PROs of children with chronic diseases. This paper aims to document the app development process and share the usability study results.

Methods

This study was conducted in Shanghai, China. This research received ethical review approval from the Second Military Medical University. The PROMIS central administration has authorized us to translate the PROMIS Pediatric and PROMIS parent proxy report short forms from English to Chinese, following the standard Functional Assessment of Chronic Illness Therapy Translation Methodology. All the translation steps and the final Chinese versions of the pediatric PROMIS short forms (C-Ped-PROMIS) have been approved by the PROMIS Statistical Center. The C-Ped-PROMIS items were semantically and conceptually equivalent to the original [6]. A total of 232 children and adolescents with cancer were interviewed with the C-Ped-PROMIS using traditional paper questionnaires. The strict and standard scale metrological test showed the C-Ped-PROMIS measures had good reliabilities and validities. Comparably, reliable and valid instrument systems can be applied in the measurement of symptoms and quality of life of children and adolescents with cancer in China, and then improve pediatric cancer care [7]. We use the validated C-Ped-PROMIS short form questionnaires in our app.

A multidisciplinary research team including health services researchers, pediatric nurses, and software engineers worked synergistically to design and develop the PROs module and administration portal. Figure 1 shows the workflow of the development process.

Group discussion: researchers, nurses and software engineers

Based on the literature review and previous experience of developing an app with a "self-assessment questionnaires" module [8-9], we conducted the first round of focus group discussion with four researchers, two nurses, and two software engineers. The discussion was conducted in a conference room at a medical university in Shanghai and lasted between 120 and 150min. All of these researchers have been involved in translating and testing C-Ped-PROMIS short forms with an educational level of a master degree or above. The nurses had more than ten years of working experience with pediatric patients and have a good understanding of C-Ped-PROMIS. The software engineers were experienced in developing medical-related software or apps.

Discussions were recorded and continued until the same themes and issues emerged from the interviewees. Discussions were transcribed verbatim within two days. Based on the functions described by the researchers, software engineers developed the

beta version of the app and web-based administration portal with basic structure, functions, and interface.

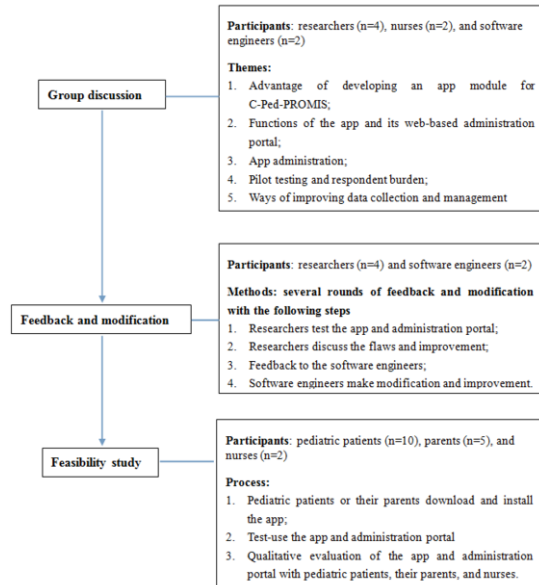


Figure 1 - The workflow of the development process

Feedback and modification: researchers and software engineers

Four researchers tested the beta version of the app individually, and then a group discussion was held to brainstorm about the disadvantages and improvement of the app. Software engineers modified the app according to the feedback. The second version of the app was developed and sent to the researchers. Another round of testing, feedback, and modification was conducted until there were no more items needed to be improved. After software engineers finished developing the web-based administration portal, researchers tested and provided feedback to the engineers for further improvement, following the same procedures as with the app's modification.

Usability study: pediatric patients, parents and nurses

Pediatric patients and parents were recruited in the Children's Hospital of Soochow University and Shanghai Children's Medical Center during August 2015. The research team introduced the app and the usability study project. No study procedures were performed until study participants signed the informed consent. The eligible pediatric patients were aged 8-17 years old with chronic diseases. The eligible parents had a child aged 5-7 years old with a chronic disease. The pediatric patients and parents were able to communicate fluently in Mandarin Chinese and had at least one Android smartphone in their family.

We first helped the pediatric patient or his/her parents and the eligible parents to install the most updated version of the app on their smartphones. The participants were asked to register and log in to fill in their basic demographic information in the app first. Then, they entered the questionnaire system to enter answers. The nurses and researchers tested the administration portal to see whether it received the accurate data users entered, and whether it was convenient to export the data for further analysis.

After the usability test, semi-structured interviews were conducted with all research participants. They expressed their

positive and negative experiences related to both the app and administration portal.

Data Analyses

Descriptive statistics were used to characterize the study participants. Two researchers transcribed the recorded qualitative data and analyzed the transcripts independently. A qualitative content analysis aimed at finding manifest, and latent meaning of data was applied to analyze the information. A stepwise approach was adopted for the content analysis. Firstly, the transcribed data was read several times by the researcher to find the theme of the whole. Secondly, the segmentation of information was done to organize the segments and sub-segments of information. Thirdly, the significant information related to research questions was extracted. Finally, data was coded and grouped into categories and abstracted into subthemes and a main theme.

Results

Basic function of the app

Six rounds of feedback and modification of the app were conducted before the final version of the app was developed. The app mainly has two modules: the demographic information module and the questionnaire module. Figure 2 (2a-2b) shows the welcome page of this app. Users need to register and log into the app. They need to fill in the demographic information and upload it as required, so they can enter the questionnaire module.



Figure 2a - Welcome page

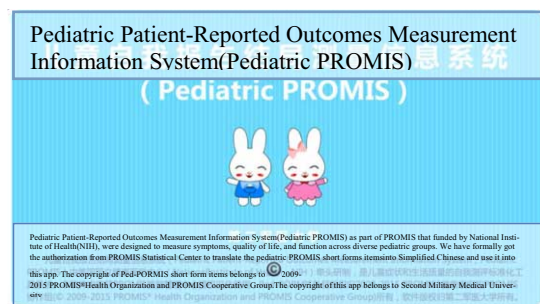


Figure 2b - English translation of welcome page

There are two entrances in the questionnaire module as is showed in Figure 3 (3a-3b). One is for pediatric patients aged from 8 to 17 years old and who are able to report by themselves, while the other entrance is for parents of pediatric patients aged

from 5 to 7 years old or pediatric patients who need help from parents to report.



Figure 3a - Two entrances of questionnaire module page



Figure 3b - English translation of two entrances of questionnaire module page

There is a voice guide, once users enter the questionnaire module, “Welcome to Pediatric PROMIS, please respond to each question or statement by choosing an item according to the feeling of the past 7 days”. The interface of the voice guide page is showed in Figure 4 (4a-4b). There are eight short questionnaire forms: Physical Functioning-Mobility, Physical Functioning-Upper Extremity, PainInterference, Fatigue, Depression symptoms, Anxiety, Peer Relationships, and Anger. If the title button of the short form is gray, the user does not need to answer that form. In other words, users can answer any of the short forms with a colored title.



Figure 4a - Voice guide page



Figure 4b - English translation of voice guide page

When all the items of one short form are answered, feedback will be provided to the user regarding the current status and care suggestions. We have added audio assistance in this app. If the pediatric patient cannot fully understand the words of any item, he/she can click the cartoon rabbit in the left corner as shown in Figure 5. The item will then be read out loud to him/her. We tried to make sure that the pronunciation and intonation of the voice have neutral emotion, so it does not affect the response of pediatric patients.



Figure 5a - Questionnaire item page

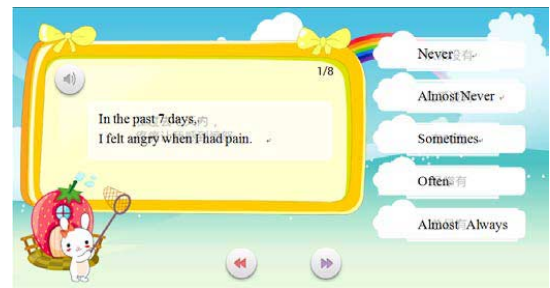


Figure 5b - English translation of questionnaire item page

Basic function of administration portal

Three rounds of feedback and modification of the administration portal were conducted before it was finalized. There are two types of administrators: the super administrator and the sub-administrator. The super administrators have access to all functions of the administration portal, while the sub-administrators do not have access to the management of administrators. Every administrator has a specific username and password. Sub-administrator's username and password are assigned by the super administrator. The super administrator has access to all of the user data, while the sub-administrator has access just to data that he/she has collected.

The administration portal lists all the demographic information that users entered. Administrators can download the information as a comma separated values file, which can be opened/imported by Microsoft Excel and almost all statistics packages. The frequency data of some demographic information can be view directly as a pie chart in the administration portal. The administrator can send specific questionnaire forms to specific participants. User ID, child's name, survey time, short form name, total scores, and total standard scores of the short form are listed in the administration system. The total standard scores are calculated according to the PROMIS Scoring Manual [10]. The administrator can also see the score of a specific item of any short form. These data can be exported to most statistical software.

The results of the usability evaluation

A total of ten pediatric patients, five parents, and two nurses participated in this usability test. The pediatric patients were aged from eight to thirteen years old. Five of them were boys.

They were diagnosed with a kind of cancer and were currently under treatment. The parents were aged from twenty one to thirty-three years old. Their children were aged from five to seven, and were diagnosed with cancer. Three of the parents were female. Two nurses were aged from 32 to 43. Both of them were female and had a college degree. Their years of working experience with pediatric patients ranged from 10 to 22 years.

After testing the app, the pediatric patients and parents felt that there was no difficulty in completing this process. All participants felt that this app was easy to use and the interfaces were friendly to children. No pediatric patients felt that the audio of the eight short forms' items affected their response to questions. The administration portal recorded all the data accurately. Nurses who tested the administration portal thought that the interfaces were simple, and it was convenient to download the data for further analysis. They spoke highly about the different ways of downloading survey results because they satisfy different aims of studies. Overall, the app and administration portal met the demand of researchers and clinical nurses. Moreover, the app was easy to use and did not add burden for pediatric patients or their parents.

Discussion

Researchers began to collect survey data through the internet years ago [11-12]. Substantial data collection efficiency, low cost, and widespread availability of internet access among diverse groups are stimulating an increased usage of web-based surveys [13]. However, while a mobile phone is now essential to people's daily lives, the research literature on using mobile apps for survey research is sparse. Although both the cost and time to program apps are high, the cost and time to collect data of a large number of participants are low [8-9]. Therefore, the app is a promising tool to collect longitudinal data because users have their own ID number and username.

The target users of this app are pediatric patients aged from 8 to 17 years old, and parents of pediatric patients aged from 5 to 7 years old. Our research team has been authorized to translate the PROMIS Pediatric and PROMIS parent proxy report short forms into Simplified Chinese. By including these two scales, this app enables researchers and practitioners to measure the quality of life and symptoms of children in different age groups. Based on the preference and interest of children, all the interfaces of the app were designed in cartoon styles. Cartoon images of a rabbit, bear, and giraffe were designed to be appealing to children. The lovely cartoon images may reduce their response burden and sense of boredom in answering the questionnaires. Some children even thought it was interesting to answer the questions because they loved the cartoon interfaces.

Due to chronic disease and treatment, some children cannot attend school like other same-age children. Some of them are undereducated to read all questions in the survey. Therefore, we added audio assistance for all items in the PRO survey. Pediatric patients usually use this app in the wards. Due to the noisy environment of wards, children tend to focus on the content of the items more than the pronunciation or intonation of the audio assistance. To some degree, the audio feature also adds interest to the survey.

The web-based administration portal collects and stores all the data users uploaded. The data can be easily downloaded in different forms for further analysis with statistical packages. The super administrators are researchers in our research team and have access to all the data. According to the authorization we received from PROMIS Health Organization, other Chinese researchers who want to use this app need to be approved by

both PROMIS Health Organization and our research team. Approved researchers will receive sub-administrator status to collect and access their own data. With our super administration role, we will be able to perform multi-site PRO research in the future.

Conclusion

This study described the process of developing a smartphone app to monitor pediatric patients' reported outcomes. The multidisciplinary research group held group discussions to clarify the app functions of the app and administration portal. Several rounds of feedback and modification were conducted to finalize the app and its administration portal. The results of the usability study showed that the app was well accepted by target users. We conclude that the App and its web-based administration portal are with good usability, and can be applied for clinical research in regards of pediatric patients' self-reported symptoms. Smartphone apps have the potential to collect survey data with high efficiency. This app is promising to promote the use of C-Ped-PROMIS in assessing the quality of life and symptoms in pediatric patients.

Acknowledgements

This study was supported by Oncology Nursing Society (ONS) Foundation and National Natural Science Foundation of China (No.71473262).

References

- [1] D. Cella, S.Yount, N. Rothrock, and et al., The patient-reported outcome measurement information system (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care* **45**(2007), S3-S11.
- [2] T.R. Walsh, D.E. Irwin, A. Meier, and et al., The use of focus groups in the development of the PROMIS pediatrics item bank. *Qual Life Res* **17**(2008), 725-735.
- [3] J.W. Varni, C.A. Limbers, and T.M. Burwinkle, Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes* **5**(2007), 2-10.
- [4] J.W. Varni, D. Thissen, B.D. Stucky, and et al., PROMIS® Parent Proxy Report Scales: an item response theory analysis of the parent proxy report item banks. *Qual Life Res* **21**(2012), 1223-1240.
- [5] J.W. Varni, D. Thissen, B.D. Stucky, and et al., PROMIS® Parent Proxy Report Scales for children ages 5-7 years: an item response theory analysis of differential item functioning across age groups. *Qual Life Res* **23**(2014), 349-361.
- [6] Y. Liu, P.S. Hinds, J. Wang, and et al., Translation and linguistic validation of the Pediatric Patient-Reported Outcomes Measurement Information System measures into simplified Chinese using cognitive interviewing methodology. *Cancer Nurs* **36**(2013), 368-376.
- [7] Y. Liu, J. Wang, P.S. Hinds, and et al., The emotional distress of children with cancer in China: an item response analysis of C-Ped-PROMIS Anxiety and Depression short forms. *Qual Life Res* **24**(2015), 1491-1501.
- [8] J. Wang, N. Yao, Y. Wang, and et al., Developing "Care Assistant": A smartphone application to support caregivers of children with acute lymphoblastic leukaemia. *J Telemed Telecare* **22**(2016), 163-171.

- [9] J. Wang, N. Yao, M. Shen, and et al., Supporting Caregivers of Children With Acute Lymphoblastic Leukemia via a Smartphone App: A Pilot Study of Usability and Effectiveness. *Comput Inform Nurs* **34**(2016), 520-527.
- [10] PROMIS Scoring Manual. Available at: <https://www.assessmentcenter.net/Manuals.aspx> (Accessed July 15, 2015)
- [11] J.C. Wyatt, When to use web-based surveys. *J Am Med Inform Assoc* **7**(2000), 426-429.
- [12] W.C. Schmidt, World-Wide Web survey research: benefits, potential problems, and solutions. *Behav Res Meth Instrum Compu* **29**(1997), 274-279.
- [13] C.A. Roster, R.D. Rogers, G. Albaum, and et al., A comparison of response characteristics from web and telephone surveys. *Int J Market Res* **46**(2004), 359-373.

Address for correspondence

Changrong Yuan. School of Nursing, Second Military Medical University, 800 Xiangyin Road, Shanghai 200433, China
Email: yuancr510@aliyun.com
Tel& Fax: +86-021-81871490

I Got 99 Problems, and eHealth Is One

Sofie Wass^a, Vivian Vimarlund^{a, b}

^a International Business School, Jönköping University, Jönköping, Sweden,

^b Department of Computer and Information Science/Human-Centered Systems, Linköping University, Linköping, Sweden

Abstract

Many eHealth initiatives are never implemented or merely end as pilot projects. Previous studies report that organisational, technical and human issues need to be properly taken into consideration if such initiatives are to be successful. The aim of this paper is to explore whether previously identified challenges within the area have remained in the Swedish eHealth setting or whether they have changed. After interviewing experts in eHealth, we present a classification of areas of concern. Recurrence of previously identified challenges was found, but also new issues were identified. The results of the study indicate that there is a need to consider organisational and semantic issues on both national and international levels. Legal and technical challenges still exist but it seems even more important to support eHealth initiatives financially, increase practitioners' knowledge in health informatics and manage new expectations from patients.

Keywords:

Health Services; Telemedicine; Medical Informatics

Introduction

eHealth is part of the field known as health informatics, which arose from the established field of medical informatics [1]. Ever since the definition of eHealth was first presented by Eysenbach in 2001, authors have provided additional definitions to capture the on-going development of this growing field. Pagliari et al. [2] draw upon the foundational definition and define eHealth as: "...the organization and delivery of health services and information using the Internet and related technologies. In a broader sense, the term characterizes not only a technical development, but also a new way of working, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide by using information and communication technology". We note the focus on the delivery of electronic health services and information to improve healthcare in various ways.

Nowadays, there is pressure to reduce costs in healthcare while the quality of service delivery to an ageing population is to be maintained [3; 4]. Healthcare organisations face several challenges, such as an increased number of elderly patients, a rise in chronic diseases and a lack of personnel [5; 6]. eHealth has emerged as a solution to the challenges, and it is expected to meet the growing demand on healthcare delivery, improve quality, expand capacity [7; 8], increase access to healthcare information and foster collaboration within and between organisations [9]. With this background, governmental and private institutions have supported the funding of eHealth initiatives in an attempt to solve these problems [10]. However, several eHealth initiatives terminate as pilot projects or are never implemented at all [11; 12]. It is estimated that 70% of eHealth projects fail to achieve their goals [13]. These failures

have been studied and described in previous studies [13; 14; 15-17]. The current literature and research in this area demonstrate that organisational change and the adjustment of work processes are crucial when eHealth solutions are to be successfully implemented [15; 16; 18-23]. A number of studies also highlight the importance of system usefulness, ease of use and flexibility, as well as individual factors, including user training, skills and user involvement [21; 24]. A model has been presented which addresses the problem of health information system (HIS) failure. This model includes dimensions related to information, technology, organisation and resources [25]. Despite this body of knowledge which has been gathered from the analysis of different healthcare systems and countries, eHealth initiatives continue to fail [11].

The aim of this study is to explore whether previously identified challenges within the eHealth area are still valid in the Swedish eHealth setting, or whether these challenges have changed. This information can be used by decision-makers as they prioritize and identify short-term and long-term goals and develop strategic plans.

Methods

A number of decision-makers and experts in eHealth were interviewed with the aim of identifying current and future challenges that the provision of eHealth in Sweden faces. The respondents included seven representatives working with the forthcoming Swedish eHealth strategy. These representatives included administrative authorities, the ministry, employers' organisations and trade organisations. They represent the broad number of actors that influence the evolution of eHealth in Sweden. The organisations that were represented were the Swedish Association of Local Authorities and Regions, the National Board of Health and Welfare, the Ministry of Health and Social Affairs, the Swedish Society of Nursing, the Swedish Association of Health Professionals, Famna and the Trade Association for the Research-based Pharmaceutical Industry in Sweden (LIF). Each respondent was contacted by email to arrange for these interviews at their offices.

The questions that were asked during the interviews covered areas such as the primary challenges and the future trends of Swedish eHealth. Background questions were also asked regarding the organisation's role in Swedish eHealth, who the important actors in the area were and significant milestones within Swedish eHealth. The interviews were performed by the first author and lasted between 40 to 65 minutes each. They were recorded and later transcribed verbatim. In addition, notes were taken during the interviews. The questions that were asked are presented in Table 1.

Table 1– Questions asked during the interviews

Could you describe the role that your organisation has played in the development of Swedish eHealth?
 Which are the most important actors, organisations or institutions within Swedish eHealth? Have they fulfilled their goals regarding eHealth?
 What are the most important milestones in Swedish eHealth?
 What are the major challenges? How are they handled?
 In which ways has Sweden reached the goals of the eHealth strategy?
 How is Sweden positioned in eHealth development, compared to other countries?
 What drives today's development of eHealth?
 What are the future trends and political questions that are discussed?
 What should Sweden prioritize in the future?
 Do you see any new challenges that organisations will face in their future work with eHealth?

Data analysis

The analysis of the interviews was performed by means of inductive content analysis, as presented by Graneheim and Lundman [26]. Content analysis is a systematic way of describing and analysing written and verbal communication [27]. The inductive approach, where themes are derived from the data, is suitable when little is known about the studied phenomena or when previous knowledge is fragmented [28]. Initially, the transcripts were read through several times to gain an overall sense of the entire data-set. Then the remarks about challenges were extracted and compiled into a section that constituted the unit of analysis.

Further to the above, the analysis was conducted in a number of steps. First, the text was divided into meaning units, i.e. "sentences or paragraphs containing aspects related to each other through their content and context" [26]. Second, each meaning unit was 'condensed'; referring to the process of shortening the meaning units while still preserving the core meaning [26]. Third, the condensed meaning units were labelled with a code. Fourth, the codes were interpreted and compared across each other so as to create mutually exclusive sub-themes. A theme can be described as "an expression of the latent content of the text" [26]. By reflecting and analysing, the initial sub-themes were reduced into fifteen sub-themes that all focused on challenges. Finally, we reflected on the sub-themes and read literature pertaining to these sub-themes and to the current state of eHealth [5; 29]. This resulted in headings that seemed to unify the sub-themes into six themes. The meaning units were read through again and nothing that contradicted the themes could be found. Even if the analysis is described as a linear process, the process was iterative and involved going back and forth during the analysis.

Trustworthiness in content analysis can be described in terms of credibility and transferability [26]. To achieve credibility, we have illustrated how sub-themes and themes are related and we present representative quotations from the interviewees that support the opinions expressed during the interviews. Transferability has been facilitated by describing the context, the data collection and the analysis process, as well as providing a rich description of the findings. In addition, the results of the analysis were discussed by the first and second author so as to ensure agreement. In the next section, we present a classification of areas of concern.

Results

From the interviews, six main areas of concern (themes) and fifteen sub-themes were identified. The respondents described issues related to different levels of interoperability, the need for more resources and placing the patient at the centre. The findings are presented in Table 2 and are described in detail with quotations from the respondents.

Table 2– Areas of concern identified from the interviews

Area of concern	Sub-theme	Respondent
Organisational interoperability	Organisational change	a, b, c, d, e, f, g
	Collaboration between organisations	a, b, c, d, e, f, g
	Common goals	a, b, e, f, g
Technical interoperability	HIS data exchange	a, c, d, f, g
	Redundancy	a, d
	High usability	a, d, e
Legal interoperability	Legislation	a, b, d, f
	Integrity	a, d, g
Semantic interoperability	Semantics	a, b, c, e
	Information structure	a, b, c, d, e, f, g
Additional resources	Increased finances	a, d, f
	Digital literacy	b, c
	Health informatics education	a, d, e, g
The patient at the centre	Patient perspective	a, c, d, e, f
	New patient expectations	a, d, f

Organisational interoperability

The importance of organisational interoperability involved managing organisational change and development, collaborating over organisational boundaries and finding common goals. The respondents described the importance of not managing eHealth initiatives as information technology (IT) projects, but, rather, as changes in business processes, structures and cultures. IT was not described as the challenging part of eHealth, compared to the organisational change that it entails. One respondent described it in the following terms: "...the technical issues are not the difficult or big problems, but changed business processes, attitudes and change management, that type is much, much bigger issues" (f).

The respondents also talked about the need to collaborate across organisational boundaries, the necessity to coordinate this collaboration and to establish spaces where actors can meet and collaborate for innovation. The same respondent said: "I think it is really great, and many people are bothered about it, they say it is so unclear in Sweden, that we have so many actors within the area of eHealth and so on. But I think that it is fantastic, I would not want another situation, it is great that it is that many. However, we might need to work a bit on coordinating different activities, collaborate to get the most out of it..." (f). The importance of having common goals was particularly stressed by the respondents. They mentioned that there is a need to have a unified picture of the future, that local ideas and goals can hamper the development of eHealth, and that everyone needs to move in the same direction. One respondent said that: "...it is not possible that one actor pushes the development. Because everyone is needed...so the difficult part, or what I see as the most important thing, is that we all move in the same pace" (b).

Technical interoperability

Another matter that was mentioned during the interviews were a number of technical challenges, including communication between different information systems and usability. The lack of communication between information systems was described as a problem that results in data redundancy, referring to the need to document the same information twice. *"It is someone that actually sits and interprets the information in the EHR [electronic health record] and transfers it manually to a web formulary"* (a). The same issue was exemplified with the use of quality registers: *"We need to establish an automated transfer of information. You document in the EHR and then it should automatically be able to transfer that information to the quality registers or other administrative systems that are for quality improvement and similar things"* (d). In addition, the respondents talked about achieving standards of 'high usability' in eHealth services and the importance of services that are supportive instead of interfering in work processes. However, technical interoperability was no longer viewed as a main barrier, compared to organisational, semantic and legal interoperability: *"...well, the technical interoperability is really the least problem today"* (a). *"The technical part is simply an element in the whole and it is not that challenging"* (f).

Legal interoperability

The respondents also mentioned the importance of legal interoperability and that laws and regulations need to adjust to accommodate the digital evolution. This also involved balancing across the fine line of integrity and service innovation. *"But it falls on the legislation, that the laws are as they are today"* (a). One respondent exemplified this point with a reference to the Swedish Data Protection Authority: *"...they only have to look at one side of the issue. As an authority, they never have to weight both scales as all others have to do. And we can say that, well maybe there is a risk for the patient, an integrity risk in one scale, but there is actually a medical risk in the other scale"* (g).

Semantic interoperability

All of the respondents stressed the importance of achieving semantic interoperability. This involved the regulation of the exchange of information between primary care and hospitals through a set of principles that includes common terms, concepts and a national information structure. One respondent described semantic interoperability as being the most important issue: *"You have this semantic interoperability with these common terms and concepts and information structure. That is such a big problem"* (a). Several respondents claimed that a unified information structure is the necessary foundation for information sharing, the development of innovative services and new ways of working.

The need for additional resources

The Swedish eHealth area was described as being in need of more resources, both in terms of financial needs, but also in terms of knowledge of how informatics and organisational healthcare issues are to be solved. Today, the various Swedish regions invest 3% of their budget on IT/IS and eHealth [30]. The respondents reported that they found it remarkable that a knowledge-intensive industry, like healthcare, does not spend more resources on eHealth. *"You need to build structures and set aside time for it. It costs money and it is an underfinanced area in total, out in the regions"* (d). In addition to the above observation, increased knowledge about health informatics was viewed as something that should be prioritized: *"...I believe that there are too few people that understand both IT quite well, understand the industry, but also understand IT so to say on a*

higher level...so you can be really good at a small part of this, but to like see all these flows" (g).

The patient at the centre

The respondents also spoke of the importance of acknowledging the patient and dealing with new expectations from citizens and patients. The healthcare sector was described as 'paternalistic' and pushing the patient into the organisational structures of healthcare. One respondent said that: *"...it is a paradigm shift within primary care, that we try to somehow adjust our healthcare organisations according to the needs of our citizens, instead of adjusting the citizens into our organisational structures"* (f). Another respondent talked about the ownership of information: *"Well you think that it is the patient that should own the information. And that trend is also a huge challenge and an important culture issue. If we succeed with it, what will the impact be then?"* (c). Patient empowerment was also mentioned as an area of concern: *"One area that we support, as an organisation, is the patients' ability to influence their own health and to gain as much knowledge as they want"* (d).

Discussion

The aim of this study was to explore whether previous reported challenges in the eHealth area are still valid in the Swedish context, or whether novel challenges can be identified. The results show that today's challenges focus on issues of interoperability, the need for additional resources and the need to manage new expectations from patients.

The most important challenge that eHealth faces seems to be organisational interoperability. The respondents stressed the importance of organisational change in relation to eHealth and the re-engineering of organisational processes. From surveying the literature, we note that individuals, tasks and technology need to fit smoothly together when eHealth services are introduced into an organisation [22; 23]. There are several reports on the importance of change management during implementation of eHealth initiatives [17; 21; 31]. Other well-known challenges, include technical interoperability that has been part of national eHealth strategies for many years [32; 33]. While our findings indicate that technology development is still a challenge, it seems to be less troublesome than before. As argued by Haux [20], this change could be identified as early as the 1990s, when organisational issues and change management became the main focus, instead of technical issues.

Further to this, certain changes in the relevant regulations and laws that govern the digital evolution, remain a barrier for the successful of eHealth. This is not a new state of affairs: legal issues constituted a barrier for the implementation of patients' digital access to EHRs in Sweden in 2003. Then, the law stated that digital access to healthcare registers could only be given to those who needed such access in order to carry out their work. Consequently, it was deemed that the patient did not need the information, since the patient was not classified as an employee, and the project was considered illegal [34]. A common information structure and shared concepts were seen as the foundation for realising the potential of eHealth. Semantic interoperability enables organisations to ensure that the exact meaning of information is understood at the moment of exchange with other actors [29]. During the last several years, a number of political incentives have been proposed in an effort to foster semantic interoperability on an international level [5; 29].

The need for additional resources was also mentioned, including both financial and knowledge resources. The respondents stressed the prerequisite of professionals that

understand how information and communication flows can be improved by eHealth. Education and training in health informatics and HIS skills have been acknowledged previously [16; 20] and discussed at MEDINFO and AMIA Symposia for several years. However, this issue seems to be less prominent in eHealth policies [35].

In summary, the challenges that were identified in the present study are well-known, and it seems as if they are the same as they were ten to fifteen years ago. However, a number of novel challenges have also emerged. When health information systems have moved on from focusing on local solutions, in, for instance an individual department, to regional or global systems [20], then new challenges seem to arise. As an example, Sweden has implemented a national patient summary as a step towards making EHRs interoperable across the different Swedish regions [36]. Today, the challenges of interoperability and organisational change seem to transcend organisational and national boundaries, and are thus not limited to the individual organisation that is to develop and implement an eHealth service. For instance, the respondents mentioned a lack of spaces where different actors can meet so that they can collaborate for innovation and share common goals. In addition, there seems to be an increasing number of new expectations from patients and citizens. In the beginning, health information systems were primarily developed to support healthcare professionals in the execution of their work [20]. But since then, several solutions have been developed to also support patients [37; 38]. Today's society is more digitalized compared to the time when eHealth was originally introduced in the 1990s, and thus the expectations from patients and citizens will be a crucial question to manage.

This study has a number of limitations that need to be taken into consideration when the reader interprets the results. First, the number of interviews was small and this limits the possibility of applying the results in other settings. An additional quantitative study could, therefore, provide further insights into the delivery of eHealth services. The respondents include representatives who work on a strategic level; future studies should also include representatives from other areas, such as healthcare professionals and entrepreneurs, for example.

Conclusions

eHealth services are globally described as an opportunity where the quality and access to healthcare can be improved and where patients can engage in understanding their own health situation [5; 6]. Previously identified challenges indicate that organisational as well as technical and human issues need to be properly dealt with if eHealth is to be successful [21; 22; 24]. Nevertheless, our study reports that these challenges still exist. Despite the public eHealth context of Sweden and the several existing eHealth policies that have been adopted in this country [39-41], the challenges that face the successful delivery of electronic health services and information to improve healthcare are similar. No strategic plan seems to be good enough, if it is not financially and administratively prioritized. Today, change management is reported as the most important issue which needs to be dealt with, along with semantic issues. These challenges are not restricted to individual organisations, but exist across organisational boundaries.

To achieve success in eHealth initiatives seems easier said than done. Fortunately, the previously identified challenges and priorities within the area of eHealth remain valid. However, the expectations that are made of eHealth services need to be realistic [42] because several problems need to be dealt with in conjunction with the development and implementation of eHealth services. If the full potential of eHealth is to be realised,

we believe that it is necessary to develop clear guidelines on both national and international levels, and that these guidelines clarify the collaboration among different actors and the implementation of eHealth services. Only then is it possible to learn from the past and develop the administrative, organisational and social structures that are needed.

Acknowledgements

The authors express their appreciation to the individuals that participated in the interviews.

References

- [1] E.J.S. Hovenga, Health Informatics - An Introduction, in: *Health informatics: an overview*, E.J.S. Hovenga, M.R. Kidd, S. Garde, and C. Hullin Lucay Cossio, eds., IOS Press, Amsterdam, 2010, pp. 9-15.
- [2] C. Pagliari, D. Sloan, P. Gregor, F. Sullivan, D. Detmer, P.J. Kahan, W. Oortwijn, and S. MacGillivray, What Is eHealth (4): A Scoping Exercise to Map the Field, *J Med Internet Res* 7 (2005), e9.
- [3] European Policy Centre, Economic governance: helping European healthcare systems to deliver better health and wealth?, in: *Coalition for Health, Ethics and Society (CHES)*, Brussels, 2013.
- [4] Business Council of Australia, Overview of megatrends in health and their implication for Australia, 2015.
- [5] European Commission, eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century, Brussels, 2012, pp. 1-14.
- [6] OECD, Health at a Glance 2013: OECD Indicators, Paris, 2013, pp. 1-211.
- [7] K.A. Stroetmann, T. Jones, A. Dobrev, and V.N. Stroetmann, eHealth is Worth it, in: *The economic benefits of implemented eHealth solutions at ten European sites*, European Communities, Luxembourg, 2006, pp. 1-59.
- [8] World Health Organization, eHealth: report by the Secretariat, in: *WHA58*, Geneva, 2005.
- [9] L. Neuhauser and G.L. Kreps, eHealth communication and behavior change: promise and performance, *Soc Semiotics* 20 (2010), 9-27.
- [10] European Commission, eHealth studies: an overview, 2013.
- [11] European Commission, Unlocking the ICT growth potential in Europe: Enabling people and businesses, Luxembourg, 2013, pp. 1-123.
- [12] R. Roine, A. Ohinmaa, and D. Hailey, Assessing telemedicine: a systematic review of the literature, *Can Med Assoc J* 165 (2001), 765-771.
- [13] J. Leviss, *HIT or miss: lessons learned from health information technology implementations*, American Health Information Management Association, Chicago, 2013.
- [14] K. Eason and P. Waterson, The implications of e-health system delivery strategies for integrated healthcare: Lessons from England, *Int J Med Inform* 82 (2013), e96-e106.
- [15] A. de Bont and R. Bal, Telemedicine in interdisciplinary work practices: On an IT system that met the criteria for success set out by its sponsors, yet failed to become part of every-day clinical routines, *BMC Med Inform Decis Making* 8 (2008), 47.
- [16] M. Lluch, Healthcare professionals' organisational barriers to health information technologies—A literature review, *Int J Med Inform* 80 (2011), 849-862.
- [17] G. Southon, C. Sauer, and K. Dampney, Lessons from a failed information systems initiative: issues for complex organisations, *Int J Med Inform* 55 (1999), 33-46.
- [18] P. Lehoux, C. Sicotte, J.L. Denis, M. Berg, and A. Lacroix, The theory of use behind telemedicine: how compatible with physicians' clinical routines?, *Soc Sci Med* 54 (2002), 889-904.
- [19] E. Balka, M. Doyle-Waters, D. Lecznarowicz, and J.M. FitzGerald, Technology, governance and patient safety: Systems issues in technology and patient safety, *Int J Med Inform* 76, Supplement 1 (2007), S35-S47.
- [20] R. Haux, Health information systems – past, present, future, *Int J Med Inform* 75 (2006), 268-281.
- [21] T. Greenhalgh, G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou, Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations, *Milbank Q* 82 (2004), 581-629.
- [22] E. Ammenwerth, C. Iller, and C. Mahler, IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study, *BMC Med Inform Decis Making* 6 (2006), 1-13.
- [23] M. Tsiknakis and A. Kouroubali, Organizational factors affecting successful adoption of innovative eHealth services: A case study employing the FITT framework, *Int J Med Inform* 78 (2009), 39-52.
- [24] M.M. Yusofa, L. Stergioulas, and J. Zugib, Health Information Systems Adoption: Findings from a Systematic Review, in: *Medinfo 2007: Proceedings of the 12th World Congress on Medical Informatics* IOS Press, 2007, pp. 262-266.

- [25] R. Heeks, Health information systems: Failure, success and improvisation, *Int J Med Inform* **75** (2006), 125-137.
- [26] U.H. Graneheim and B. Lundman, Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness, *Nurs Educ Today* **24** (2004), 105-112.
- [27] K. Krippendorff, *Content analysis : an introduction to its methodology*, Sage Thousand Oaks, Calif., 2004.
- [28] S. Elo and H. Kyngäs, The qualitative content analysis process, *J Adv Nurs* **62** (2008), 107-115.
- [29] European Commission, European Interoperability Framework (EIF) for European public services, 2010, pp. 1-35.
- [30] L. Jerlval and T. Pehrsson, eHealth in the county councils [in swedish], 2016, pp. 1-26.
- [31] V. Vimarlund, N.-G. Olive, I. Scandurra, and S. Koch, Organizational effects of information and communication technology (ICT) in elderly homecare: a case study, *Health Inform J* **14** (2008), 195-210.
- [32] C. Chronaki, A. Estelrich, G. Cangiali, M. Melgara, D. Kalra, Z. Gonzaga, L. Garber, E. Blechman, J. Ferguson, and S. Kay, Interoperability standards enabling cross-border patient summary exchange, in: *Medinfo 2014: Proceedings of the 14th World Congress on Medical and Health*, IOS Press, 2014, pp. 256-260.
- [33] A. Moen, W.O. Hackl, J. Hofdijk, L. Van Gemert-Pijnen, E. Ammenwerth, P. Nykanen, and A. Hoerbst, eHealth in Europe - status and challenges, *Eur J Biomed Inform* **8** (2012), 2-7.
- [34] G. Erlingsdóttir and C. Lindholm, When patient empowerment encounters professional autonomy: The conflict and negotiation process of inscribing an eHealth service, *Scand J Public Admin* **19** (2015), 27-48.
- [35] S. Khoja, H. Durrani, P. Nayani, and A. Fahim, Scope of Policy Issues in eHealth: Results From a Structured Literature Review, *J Med Internet Res* **14** (2012), e34.
- [36] E.C. Lehnborn, A.J. McLachlan, and J.-a.E. Brien, A Qualitative Study of Swedes' Opinions about Shared Electronic Health Records, in: *Medinfo 2013: Proceedings of the 14th World Congress on Medical and Health*, IOS Press, 2013, pp. 3-7.
- [37] M.J. Ball and J. Lillis, E-health: transforming the physician/patient relationship, *Int J Med Inform* **61** (2001), 1-10.
- [38] D. Masys, D. Baker, A. Butros, and K.E. Cowles, Giving Patients Access to Their Medical Records via the Internet: The PCASSO Experience, *J Am Med Inform Assoc* **9** (2002), 181-191.
- [39] Ministry of Health and Social Affairs, National IT strategy for health and social care [In Swedish], Printfabriken, Karlskrona, 2006, pp. 1-46.
- [40] Ministry of Health and Social Affairs, National eHealth - the strategy for accessible and secure information in health and social care [In Swedish], Edita, Västerås, 2010, pp. 1-40.
- [41] Ministry of Health and Social Affairs and Swedish Association of Local Authorities and Regions, Vision eHealth 2025- a common starting-point for digitalization in health and social care [In Swedish], 2016, pp. 1-16.
- [42] J. van Gemert-Pijnen, N. Nijland, M. van Limburg, H.C Ossebaard, S.M Kelders, G. Eysenbach, and E.R Seydel, A holistic framework to improve the uptake and impact of eHealth technologies, *J Med Internet Res* **13** (2011), e111.

Address for correspondence

Sofie Wass, sofie.wass@ju.se +46 (0)36-101790

iT2DM Project: A Framework for Secondary Use of EHR Data for High-Throughput Phenotyping in Diabetes

Yufang Wei^a, Yujuan Shang^a, Jie Wang^a, Defu He^a, Shengrong Zhou^a, Weixun Zhang^a,
Lili Shi^a, Yalan Chen^a, Kui Jiang^a, Huiqun Wu^{a,b}, Jiancheng Dong^a

^aDepartment of Medical Informatics, Medical School of Nantong University, Nantong, China

^bDepartment of Biomedical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, USA

Abstract

Diabetes is one of the major burdens in health care, but could be controlled if the relevant data are well-managed. Referring to current successful cases, we designed a framework for the interoperability and integration of medical data in compliance with both archetype and reference information model specification. The clinical data model (CDM) was designed on the basis of OpenEHR archetypes and self-made patient generated health data (PGHD). Integrating healthcare enterprise (IHE) protocol was taken into integrating different modality data. After terminology mapping, the personal health record could be transferred and shared in different clinical information vendors complying with HL7 standards. Many fragment data such as blood glucose and gene data were also integrated to system. Those patients suspected of higher risk of diabetic retinopathy (DR) were grouped as case and other patients could be filtered as control cohort. Furthermore, the framework could be further developed for precision medicine.

Keywords:

Diabetes Mellitus; Health Level Seven; Precision Medicine

Introduction

Diabetes, a chronic metabolic disorder of blood sugar, cause a variety of complications including heart disease, stroke, chronic renal failure, foot ulcers and retinopathy, if blood sugar is not be controlled at normal level. By 2015, there were about 415 million people suffering from diabetes in the world, of which 90% are type 2 diabetes. Between 2012 and 2015, about 1.5 to 5 million people worldwide died from diabetes each year, resulting in direct economic losses of about \$ 612 billion[1-3]. With rapid development of high-throughput biotechnology, various types of omics and systematic methods, precision medicine is becoming a new paradigm of future medical development. As the next generation of diagnosis and treatment technology, precision medicine has important theoretical and practical significance.

Precision medicine is a biomedical strategy of health care to improve patient-specific and individualized diagnoses, medical decisions, medications, therapies, and prognoses and to make the life quality better[4-6]. For diabetes management, precision medicine services based on big data will allow for patients to be served with appropriate treatment at the right time. Therefore, effective secondary use of relevant health data is a critical step. In recent years, hospitals, patients and doctors have paid more and more attention to personal health management. The opening and sharing of medical data has

become a hot topic in the world. The United States “Blue Button” took the lead in achieving the opening of personal medical data. Users can download their own medical data online, master more personal health information and family medical information. According to statistics, more than 107 million people use “Blue Button”, more than 250 million users registered for Blue Button, more than 5.2 million “Blue Button” files were downloaded, and more than 84 million patients choose to use Secure Messaging. In addition to Blue Button, the United States Partners HealthCare Group (PHG) health service system can also provide patients with records integrated with disease, medications and test results. The OpenNotes project allows patients to access physician diagnostic notes online.

There are many diabetes information management systems (DIMS). However, existing DIMS with standalone data repository has the limitation of interoperability. In this paper, we propose a new diabetes framework using the clinical data model (CDM)[7-10] and following health level 7 (HL7) standards, an integrating with data from third-party systems such as blood glucose, fundus image, etc. Additionally, cohort grouping were analyzed and developed for precision medicine.

Methods and Results

System design and implementation

The development environment of our study were as follows: CPU: P4 1.8GHz; RAM: 256MB; Operating system: Windows 10; Database: SQL Server 2000; Server: Local server. The following program was required to run the website normally: Microsoft Visual Studio 2013; Runtime environment: CPU: Intel Pentium M 1.88GHz; RAM: 1GB; Hard disk: 60GB. The system was designed as browser/server (B/S) structure and with a simple portal interface for user. After successful registration, users can login into the website and upload their health data by manual entry or direct file upload. The ability to download data as a separate file for purposes of sharing and integration with other information vendors were also developed. (Figure 1)

Phenotyping, querying and grouping

Phenotype is key to big data within a platform achieving precision medicine[19]. Identification of patient cohorts for conducting clinical and research studies has always been a major bottleneck and time-consuming process[20]. To meet this requirement, increasing attention is being paid to leverage electronic health record (EHR) data for cohort identification[21]. In particular, with the increasing adoption of EHRs for routine clinical care within the U.S, as a result to Meaningful Use, evaluating the strengths and limitations for secondary use of EHR data has important implications for clinical and translational research, including clinical trials, observational cohorts, outcomes research, and comparative effectiveness research.

In the recent past, several projects, including eMERGE, SHARPN, and i2b2, have developed tools and technologies for identifying patient cohorts using EHRs. A key aspect of this process is to define inclusion and exclusion criteria involving EHR data fields such as diagnoses, procedures, laboratory results, medications as well as logical operators. Therefore, we developed a “query and grouping” module and scalable informatics infrastructure for normalization of both structured and unstructured EHR data into a unified, concept-based model for high-throughput phenotype extraction. The phenotyping algorithms were developed by the domain experts and clinicians along with informaticians. After de-identification of patients’ privacy information, executable Drools rules scripts and the rules engine that executes the scripts were used to generate evaluation results and visualization. Domain experts and IT personnel performed extract-transform-load (ETL) of “native” clinical data from the EHR systems into a normalized and standardized CDM database that was then accessed by the JBOSS rules engine for execution (Figure 6).

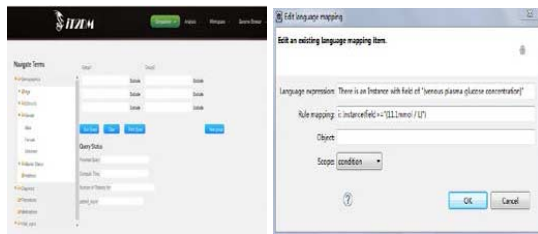


Figure 6 - The interface of querying and grouping (left) and phenotyping (right)

Blood glucose data integration based on HL7 V2 message

In this experiment, Mirth Connect was utilized as an interface engine for HL7 data transmission and integration. To validate the HL7 message, 7edit was used to validate and parse HL7 message in this test (Figure 7).

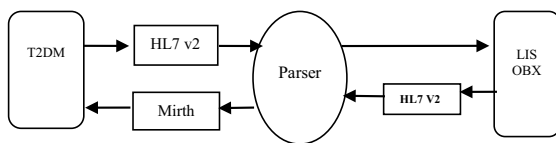


Figure 7 - An integrated flow chart of HL7 v2 messages in OpenEMR

In this study, a blood test was ordered from the system as a request and sent to the laboratory information system (LIS) as a new order waiting for glucose value fulfillment (Figure 8).



Figure 8 - The blood glucose examination doctor's orders

After order automation, glucometer produces the results of the examination including series number, result type, result unit, normal reference range, etc. The glucose value in OBX segment of HL7 message were transferred with HTTP Listener, and a novel channel was built up by Mirth to configure the corresponding transmission channel in the mirth and insert the blood glucose data into the procedure_report table through the SQL statement and the integrated result could be shown (Figure 9).

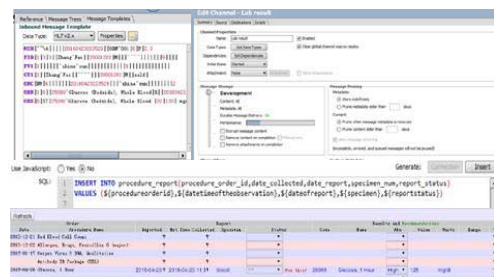


Figure 9 - Some shot screens of our integration experiments

Gene reports sharing and integration based on IHE-LAB

IHE-LAB specification is a framework to carry out the integration between laboratory examination and EMR. According to CDA specification, we edited relevant procedure with the help of XML parser to conduct any other operation or save medical records into local database. CDA has an extensive range of vocabulary, also includes a number of other organization vocabulary, which is recognized by the HL7 specification, such as logical observation identifiers names and codes (LOINC), systematized nomenclature of medicine - clinical terms (SNOMED-CT), etc. HL7 reference information model (RIM) is the model basis of HL7 V3 standards which formed on the CDA document stated. XML technologies inherited implementation of the HL7 specification (ITS), which produces CDA data types. On the other hand, CDA does not consider transmittal modes that happen to be a part of HL7 standards. Regarding clinical documents, CDA specifications complement the HL7 standard. In this research, a diabetic retinopathy gene test report (DR-GTR) in compliance with CDA format was created and such GTR was integrated in the form of continuity of care record (CCR) into our system (Figure 10).



Figure 10 - DR GTR integrated and visualized

Fundus image integration based on IHE-XDS

Portable fundus camera (Suzhou KangJie Medical Equipment Co., Ltd, China) was used to obtain fundus images. The raw jpg format of images were transformed into digital imaging and communication in medicine (DICOM) format, then transmitted into our previously established eye-PACS[22]. Users could view fundus images by visiting web accessible web access to DICOM persistent objects (WADO) protocol within the system (Figure 11).

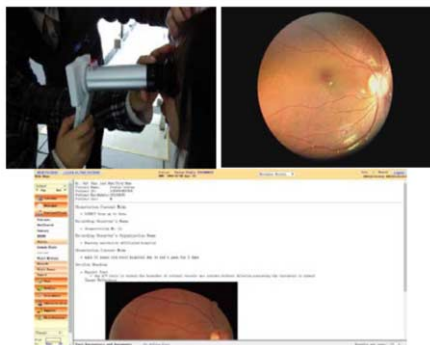


Figure 11 - Fundus images were integrated and shown (up: fundus image obtained; down: fundus image integrated)

Discussion

Our team has previously developed an EHR system for local Nantong residents with about 6 millions records. However, as more and more fragment data from diabetic patients are now being obtained from different information vendors, a centralized data repository is not flexible enough to afford increasing amounts of fragment data. The management of such data from all aspects is therefore in need of fully integration of different standalone vendors. Many institutes have launched studies on such sharing platform developments. The CDA standard that specifies the structure and semantics of clinical documents used for exchange was adopted for integration. Due to the ability of clinical descriptions, flexibility and machine readable, CDA template was implemented as data specification for other systems[17; 23; 24]. However, archetypes based model have been proved to be more flexible for system architecture and have been taken into practice in many health care system development[25; 26]. The OpenEHR specification is an open EHR architecture, which is developed by the OpenEHR institutions, and its goal is to achieve the sharing of medical information within

systems[27]. The two-layer model is designed to loosely couple the medical model and software engineering[18; 28; 29]. Our system adopts such two-layer model and collects, opens, shares and analyzes patient's health care data to help patients with diabetes monitor their all-weather, full life-cycle of health indicators. The value of the framework is not only that people with diabetes can obtain data from EHRs, they can also share and link their other modalities of data into EHRs, thus contributing to faster and more accurate management of diabetes.

Since precision medicine has been proposed, a number of research projects have been granted to initiate cohort studies with large sample size and abundant health data, which stimulates the need for development of high-throughput data systems and electronic platforms. SHARPN, i2b2 and transSMART are such influential open-source platforms that could be integrated with genomic data and are important for clinical trials investigations, and meaningful to precision medicine. Similarly, we developed the "query and grouping" module referring to i2b2 and SHARPN project. Moreover, we made the EHR data standardization via mapping and interface for transforming, that is, uniform terminology was used to facilitate the researchers who choose concepts to query on. By doing so, clinicians could start grouping and cohort research quickly and easily. The large-scale precision medicine investigation relies on phenotyping algorithms, and there have been two challenges in particular, that is, local data elements in an EHR may not be natively represented in a format consistent with the QDM including the required code systems and value sets; and an EHR typically does not natively have the capability to automatically consume and execute phenotype definitions logic. In this study, we tried to map local data elements to SNOMED or ICD codes to solve term definition disambiguation.

Based on our framework, patients with diabetes could obtain their own medical records to effectively carry out self-health management and improve the quality of medical treatment. Doctors could provide personalized care to patients with diabetes, reducing the risk of misdiagnosis, thus saving medical costs and increasing benefits. The innovation in this study relies on PGHD CDM, since it could provide potential solution to enabling prospective cohort design by altering tests and to patients with feedback reports. Although our framework can effectively integrate data in heterogeneous information systems, including the integration of clinical data, we also need to consider the information integrity and confidential transmission and other security issues. Additionally, we developed some visualization plugins to allow clinicians to better analyze diabetic data. The ultimate goal is to achieve a framework for secondary use of EHR data for high-throughput phenotyping in diabetes. Furthermore, we will develop the computer aided decision analysis module[30] based on our diabetes platform to carry out the trial and investigation in the community, and improve our system via the study of the feedback information.

Conclusion

In conclusion, iT2DM could facilitate data sharing and integration of diabetic patients, and could be further developed for precision medicine specified on diabetes management.

Acknowledgements

This work was supported by the grant of National Natural Science Foundation of China (No. 81501559, 81271668, 61671255), Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (No.15KJB310015, 14KJB510031) and Science and Technology Project of Nantong City (MS12015105), Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-aged Teachers and Presidents 2016, Science and Technology Project Nantong University (15Z02, 15Z04), Graduate Research and Innovation Plan Project of Nantong University (YKC15056, YKC16072). We would like to thank Suzhou Kangjie Medical Inc Co, Ltd. for their support of digital fundus camera for this study.

References

- [1] Y. Shi and F.B. Hu, The global implications of diabetes and cancer, *Lancet* **383** (2014), 1947-1948.
- [2] Z. Tao, A. Shi, and J. Zhao, Epidemiological Perspectives of Diabetes, *Cell Biochemistry and Biophysics* **73** (2015), 1-5.
- [3] I. Vlad and A.R. Popa, Epidemiology of Diabetes Mellitus: A Current Review, *Romanian Journal of Diabetes Nutrition and Metabolic Diseases* **19** (2016), 433-440.
- [4] J.S. de Bono, A New Initiative on Precision Medicine, *New England Journal of Medicine* **372** (2015), 793-795.
- [5] T.R. Insel, The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry, *American Journal of Psychiatry* **171** (2014), 395-397.
- [6] J.L. Jameson and D.L. Longo, Precision medicine—personalized, problematic, and promising, *New England Journal of Medicine* **372** (2015), 2229-2234.
- [7] P.R. Silva-Ferreira, J.H. Patriarca-Almeida, P.M. Vieira-Marques, and R.J. Cruz-Correia, Improving expressiveness of agents using openEHR to retrieve multi-institutional health data: Feeding local repositories through HL7 based providers, in: *Information Systems and Technologies*, 2012, pp. 1-5.
- [8] E. Sundvall, M. Nyström, H. Petersson, and H. Ahlfeldt, Interactive visualization and navigation of complex terminology systems, exemplified by SNOMED CT, *Studies in Health Technology & Informatics* **124** (2006), 851-856.
- [9] E. Sundvall, R. Qamar, M. Nyström, M. Forss, H. Petersson, D. Karlsson, H. Ahlfeldt, and A. Rector, Integration of tools for binding archetypes to SNOMED CT, *BMC Medical Informatics and Decision Making* **8 Suppl 1** (2008), 1-10.
- [10] T. Vianteeravatt, M.N. Anyanwu, V.R. Nagisetty, E. Kusu, M.E. Sakauye, and D. Wu, Clinical data integration of distributed data sources using Health Level Seven (HL7) v3-RIM mapping, *Journal of Clinical Bioinformatics* **1** (2011), 32.
- [11] G.P. Gujarathi and Y.S. Ma, Parametric CAD/CAE integration using a common data model, *Journal of Manufacturing Systems* **30** (2011), 118-132.
- [12] R. Lfleur, L. Hcourt, R. Mcaliff, Richard Platt, J. Vselby, and J. Sbrown, Launching PCORnet, a national patient-centered clinical research network, *Journal of the American Medical Informatics Association* **21** (2014), 578-582.
- [13] T.A. Oniki, N. Zhuo, C.E. Beebe, H. Liu, J.F. Coyle, C.G. Parker, H.R. Solbrig, K. Marchant, V.C. Kaggal, and C.G. Chute, Clinical element models in the SHARPN consortium, *Journal of the American Medical Informatics Association* **23** (2015), 248-256.
- [14] X. Zhou, S. Murugesan, H. Bhullar, Q. Liu, B. Cai, C. Wentworth, and A. Bate, An evaluation of the THIN database in the OMOP Common Data Model for active drug safety surveillance, *Drug Safety* **36** (2013), 119-134.
- [15] R.N. Agbara, Promoting Patient Engagement in Meaningful Use of the EHR, *Dissertations & Theses - Gradworks* (2014).
- [16] C.M. Costa, M. Menárguez-Tortosa, and J.T. Fernández-Breis, Clinical data interoperability based on archetype transformation, *Journal of Biomedical Informatics* **44** (2011), 869-880.
- [17] D. Moner, A. Moreno, J.A. Maldonado, M. Robles, and C. Parra, Using archetypes for defining CDA templates, *Studies in Health Technology & Informatics* **180** (2012), 53-57.
- [18] D.W. Van, Meaningful use of patient-generated data in EHRs, *Journal of Ahima* **84** (2013), 30-35.
- [19] P.N. Robinson, Deep phenotyping for precision medicine, *Human Mutation* **33** (2012), 777-780.
- [20] D. Li, C.M. Endle, S. Murthy, C. Stancel, D. Suesse, D. Sottara, S.M. Huff, C.G. Chute, and J. Pathak, Modeling and executing electronic health records driven phenotyping algorithms using the NQF Quality Data Model and JBoss® Drools Engine, *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* **2012** (2012), 532-541.
- [21] J. Pathak, K.R. Bailey, C.E. Beebe, S. Bethard, D.C. Carrell, J.C. Pei, D. Dligach, C.M. Endle, L.A. Hart, and P.J. Haug, Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPN consortium, *Journal of the American Medical Informatics Association* **20** (2013), 341-348.
- [22] H.Q. Wu, Z.M. Lv, X.Y. Geng, K. Jiang, L.M. Tang, G.M. Zhou, and J.C. Dong, Interoperative fundus image and report sharing in compliance with integrating the healthcare enterprise conformance and web access to digital imaging and communication in medicine persistent object protocol, *International Journal of Ophthalmology* **6** (2013), 879-883.
- [23] R.H. Dolin, B. Rogers, and C. Jaffe, Health level seven interoperability strategy: big data, incrementally structured, *Methods of Information in Medicine* **54** (2014), 75-82.
- [24] W. Goossen and L.H. Langford, Exchanging care records using HL7 V3 care provision messages, *Journal of the American Medical Informatics Association* **21** (2014), 363-368.
- [25] H. Leslie, International developments in openEHR archetypes and templates, *Health Information Management Journal* **37** (2008), 38-39.
- [26] L. Wang, L. Min, R. Wang, X. Lu, and H. Duan, Archetype relational mapping - a practical openEHR persistence solution, *BMC Medical Informatics and Decision Making* **15** (2015), 1-18.
- [27] Mart, C. Nez-Costa, Men, M. Rguez-Tortosa, Fern, and J.T. ndez-Breis, An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes, *Journal of Biomedical Informatics* **43** (2010), 736-746.
- [28] G.F. Cooper, B.G. Buchanan, M. Kayaalp, M. Saul, and J.K. Vries, Using computer modeling to help identify patient subgroups in clinical data repositories, *Proceedings / AMIA ... Annual Symposium. AMIA Symposium* **5** (1997), 180-184.
- [29] B. Haarbrandt, E. Tute, and M. Marschollek, Automated Population of an i2b2 Clinical Data Warehouse from an openEHR-based Data Repository, *Journal of Biomedical Informatics* **63** (2016), 277-294.
- [30] A. González-Ferrer, M. Peleg, B. Verhees, J.M. Verlinden, and C. Marcos, *Data Integration for Clinical Decision Support Based on openEHR Archetypes and HL7 Virtual Medical Record*, Springer Berlin Heidelberg, 2013.

Address for correspondence

Kui Jiang: kuij@ntu.edu.cn

Huiqun Wu: wuhuiqun@ntu.edu.cn

Development of Home-Based Sleep Monitoring System for Obstructive Sleep Apnea

Peirong Wu^a, Guan-Ting Chen^a, Yanyan Cui^a, Jin-Wei Li^a, Terry B.J. Kuo^b, Polun Chang^a

^a Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

^b Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan, China

Abstract

Obstructive Sleep Apnea (OSA) has been proven to increase the risk of high blood pressure, heart attack, stroke, obesity, and diabetes. If people would like to know whether they are suffering from this sleep disorder, they need to go to particular hospital with which a sleep center that could perform polysomnography (PSG); however, for most people, this is not convenient. Consequently, the goal of this study is to develop a convenient, lower priced, and easy-to-use home-based sleep monitoring system. The researchers have developed the "Sleep Healthcare Management System" for OSA patients and healthcare providers. It combines smartphone and wearable devices that can perform real-time sleep monitoring. Healthcare providers could apply their professional knowledge to provide customized feedback via a web application. When the patient is diagnosed with an abnormal sleep health condition, healthcare providers may be able to provide appropriate and timely care.

Keywords:

Polysomnography; Smartphone; Obstructive Sleep Apnea

Introduction

Obstructive Sleep Apnea (OSA) is a common disorder in the population [1]. The National Heart, Lung and Blood Institute of American National Institutes of Health (NIH) indicated that untreated sleep apnea could increase the risk of hypertension, heart attack, stroke, obesity, and diabetes, etc.

OSA is one type of sleep apnea. According to different symptoms, there are three types of sleep apnea: OSA, central sleep apnea (CSA), and mixed sleep apnea. Among these types of sleep apnea, OSA is the most common. The prevalence of OSA was a mean of 22% in men and 17% in women [1].

The standard test for the diagnosis is polysomnography (PSG). This test would monitor multiple physiological signals of sleeping patients overnight in order to measure the apnea-

hypopnea index (AHI). According to the American Academy of Sleep Medicine (AASM), the diagnosis of OSA is $AHI \geq 5$ with symptoms such as daytime somnolence, fatigue, insomnia, mood disorders, and cognitive impairment, or cardiovascular comorbid conditions [2].

However, PSG is not easily accessible in Taiwan, people need to go to the sleep center to do the examination. The subject must be settled with many sensors for collecting physiological signals during the test, the subject often feels uncomfortable; furthermore, the subject usually has only one night to do this test, PSG may not be suitable for those particular sleep disorders which should take more time to be tracked.

Based on the above reasons, the researchers want to develop a convenient, low price and easy-to-use home-based system. Thus, people could realize their sleep health condition in an easy way.

Methods

The researchers designed a home-based sleep monitoring system, the "Sleep Healthcare Management System" [3-7]. This system includes two subsystems: A front-end subsystem "Sleep Healthcare Management System-Patient" and a back-end subsystem "Sleep Healthcare Management System-Healthcare Provider". At the front-end subsystem, we would develop a mobile phone application for patients to collect their physiological signals and view these physiological signals in a visualized way. At the back-end subsystem, we would develop a real-time feedback system, which is a web application for healthcare providers to view physiological signals of the patients and set the content of feedback that would be sent to the patients. The entire system structure is shown in Figure 1. All functional requirements of the front-end subsystem and the back-end subsystem are shown in Table 1 and Table 2 respectively.

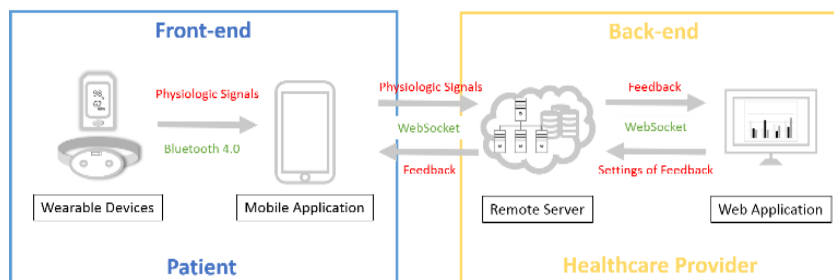


Figure 1 – System Structure: Sleep Healthcare Management System

Table 1 – Functional Requirements: Sleep Healthcare Management System-Patient.

Functionality	Requirements
Dynamic Scale	fill out the scale; add extra explanation of question; different types of question(e.g. single choice, multiple choice, numeric, short answer, photo, date); answer temporary saving
Sleep-Position Trend	display chart(e.g. sleep-position, heart rate, SpO2 trends); display chart in different time interval(e.g. 1 minute, 10 minutes, 30 minutes, 1 hour); display chart in different date range; export data; total time of every sleep-position; statistical distributions of SpO2(e.g. interval 100-98, 97-95, 94-90, 89-80, <80)
Sleep Management	show sleep parameters(e.g. total sleep time, total light sleep time, total deep sleep time, total wake up time, turnover times, wake up times); display the chart of the sleep stage
SpO2 Device	show real-time physiological signals (e.g. SpO2, PR, PI)
Breath Tracker	display the real-time of physiological signals and chart(e.g. ECG, heart rate trends, breath frequency, breath time)
Device Management	wearable device binding; the mobile phone should receive physiological signal from wearable devices continually at night
Data Collection	system preferences displaying(e.g. whether support Bluetooth technology; whether Bluetooth is enabled; wearable device connecting status; store the data received from wearable device)
Register & Login	account registry; system login

message, health education); show feedback scope(e.g. type, name, content)

Physiological Signals Feedback

logical judgment; logical rule(ex: greater than, smaller than, equal to, between); send feedback(e.g. instant response, warning message)

Register & Login

account registry; system login

Evaluation of the Sleep Healthcare Management System

In this system, the sleep-position trend is the most important reference of the sleep diagnosis. The accuracy of the sleep-position detection algorithm reflects the reliability of this system. Therefore, the researchers designed a controlled trial followed by a “sleep-position-detection evaluation procedure” (Figure 2) to evaluate our system.

In this controlled trial, all recruited subjects will be assigned to the experiment group.

During the trial, the control group would use a timer to record the moment of every sleep-position changes; the experiment group would wear the breath tracker and follow the “sleep-position-detection evaluation procedure” to change their sleep-position.

After that, according to the Karl Pearson's Coefficient of Correlation, we could calculate the accuracy of the sleep-position detection and the coefficient of correlation of the sleep-position detection.

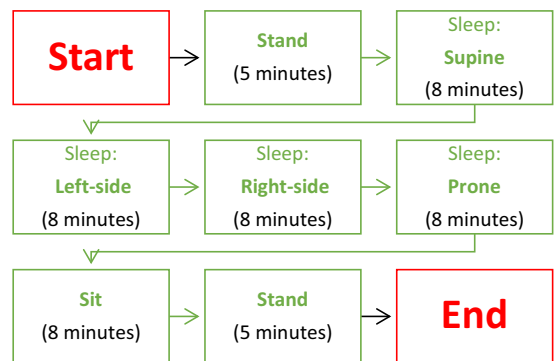


Figure 2 – Sleep-Position-Detection Evaluation Procedure.

Table 2 – Functional Requirements: Sleep Healthcare Management System-Healthcare Provider.

Functionality	Requirements
Sleep-Position Trend	display the chart of the trend of physiological signals(e.g. sleep-position, heart rate, SpO2); display the chart of the trend of physiological signals in different time interval (e.g. 1 minute, 10 minutes, 30 minutes, 1 hour); display the chart in different range of date; exporting data; total time of every sleep-position; draw the statistical distributions of SpO2(e.g. intervals: 100-98, 97-95, 94-90, 89-80, <80)
Feedback Setting	different types of feedback(e.g. instant response, scale, warning

Results

According to the system design and all functional requirements, we developed a home-based sleep monitoring system called “Sleep Healthcare Management System”. This system includes two subsystems: “Sleep Healthcare Management System-Client”, which is a mobile phone application and “Sleep Healthcare Management System-Healthcare Provider”, which is a web application.

Patients could view their physiological signals which were collected by the SpO2 device and breath tracker in a visualized way via their mobile phone; the healthcare provider could view multiple physiological signals which were uploaded by patients on a web application. The detail of these two subsystems will be described below:

Front-end: Sleep Healthcare Management System-Patient

Android is the most popular mobile OS and it provides a free, open source, cross-platform development environment. Besides, numerous open source projects in Android support arithmetical and graphical functions. Also, the developers could use its official development tool: Android Studio to develop applications in a convenient way [9].

Therefore, we adopted Android as the mobile OS of this mobile phone application and Android Studio 2.0 as the development tool; we run this application on mobile phone, which supported Android version 4.3 or later.

This mobile phone application contains 7 modules: dynamic scale, sleep management, sleep-position trend, data collection, SpO2 device, device management, breath tracker (Figure 3). Among these 7 modules, SpO2 device module and breath tracker module need to manipulate with wearable devices for collecting the physiological signals of the patient.

Sleep-Position Trend Module

This module is used for presenting the change of the patient's sleeping position (e.g. supine, right-side, left-side, prone, stand) in a visualized way. The healthcare provider could view charts in different time segments (e.g. 1 minute, 10 minutes, 30 minutes, 1 hour) and export these chart (Figure 5).

During 8 hours sleep, most people would turnover up to dozens of times, even hundreds of times. Appropriate turn-over times is a self-protection mechanism but turnover too many times might lower the quality of sleep. Because of turnover the deep sleep time could be reduced. So the times of turnover is an important reference index to health professionals. Furthermore, the variation of SpO2, heart rate and sleep-position are the important to health professionals, too.

Sleep Management Module

This module is used for presenting the change of the patient's sleeping parameters (e.g. sleep start time, sleep end time, sleep total time, light sleep time, deep sleep time, wake up time, wake up times and sleep score) in a visualized way.

All sleeping parameters are obtained from algorithm [8] that process physiological signals transmitted from wearable devices to the data collection module of this system via Bluetooth technology.

SpO2 Device Module

This module could present real-time physiological signals of the patient: SpO2, pulse rate(PR), perfusion index (PI). In the normal condition, pulse rate equals heart rate. PI could reflect the situation of patient's limb perfusion in a visualized way.

Breath Tracker Module

This module could present real-time physiological signals of HRV report of the patient: LF(low frequency), HF(high frequency), total power(TP), RR(R-R interval), RT(reserve time), IT(inspiratory time), IV(inspiratory volume).

Device Management Module

Patients could bind two wearable devices: SpO2 device and breath tracker, to their mobile phone via this module.

Data Collection Module

The main function of this module not only shows real-time physiological signals collected from two wearable devices on the patient's mobile phone, but also shows the connection status between wearable devices and the patient's mobile phone.

Dynamic Scale Module

In our system, we regard scale(questionnaire) as one type of feedback. The healthcare provider could view physiological signal charts and learn more about the sleep health condition of the patient by using scale as an evaluating tool. Because the

content of scale would be a variety of forms, we separate scale as an individual module. Content of the scale was designed by healthcare professionals and saved in the remote server.

As the healthcare provider is conscious of the patient, who is under a bad sleep health condition, the scale might be sent to the patient, then healthcare provider could evaluate the sleep health condition of the patient via scale.

There are two reasons for why we called this module "dynamic" scale module,: First, according to the sleep health condition of the patient at different time, the scale which this system decided to send would change dynamically. Second, when the patient finished a scale, the system would calculate a score of the finished scale for judging whether it is necessary to have another scale for evaluating the patient's sleep health condition or not; if the score meets the scale-sending criteria, this system will send another scale to the patient.

Thus, this "Dynamic Scale" would work dynamically, it could send scale, which meets different sleep health conditions of the patient dynamically.

Back-end: Sleep Healthcare Management System-Healthcare provider

One of the main function of this web app is to perform the real-time change of the patient's physiological signals via charts. Consequently, the efficacy of this web application becomes the most important consideration in development.

Node.js is a web application framework built on Chrome's V8 JavaScript engine. Because of its' lightweight and efficiency, Node.js becomes a popular programming language for web development now; moreover, we could use Node.js' open source library: NPM for trimming development time greatly [10].

Because Node.js exactly met our requirement, we adopted Node.js as the main program language of this web application and WebStorm as the development tool.

WebStorm is a lightweight yet powerful IDE, equipped for complex client-side development and server-side development with Node.js.

This web application contains 3 modules: sleep-position trend, physiological signals feedback, feedback setting (Figure 4).

Sleep-Position Trend Module

This module is used for presenting the change of the patient's sleeping position(e.g. supine, right-side, left-side, prone, stand). Healthcare providers could view charts in different time segments(e.g. 1 minute, 10 minutes, 30 minutes, 1 hour) and export these chart(Figure 6).

Physiological Signals Feedback Module

This module is an association-rule system, healthcare providers could record clinical principles or their professional knowledge as rules(ex: greater than, smaller than, equals, between) via web application, all rules would be stored in the remote server.

According to patient's physiological signals(e.g. heart rate, sleep-position, temperature, SpO2, sleep score), these rules are used for analyzing the sleep health condition of the patient.

If the result of analysis shows that the patient is under the abnormal sleep health condition, or the patient's sleep health condition conforms to the rule, the feedback would be sent to the patient.

Feedback Setting Module

Healthcare providers could decide what content of feedback would be sent to patients via this module. The content of feedback would be many formats, such as text message, questionnaire and health education information. Healthcare

providers could use questionnaires to help diagnosis and know more about the sleep health condition of the patient.

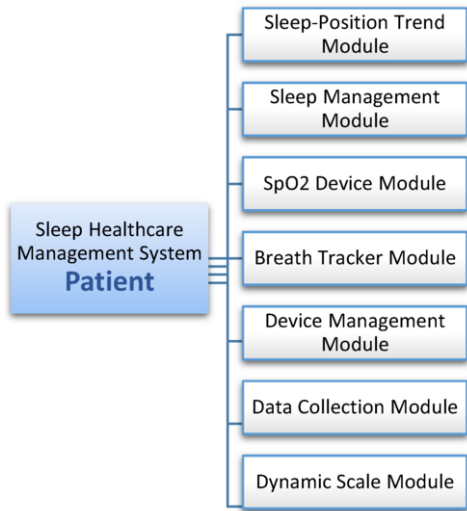


Figure 3 – System Structure: Sleep Healthcare Management System-Patient.

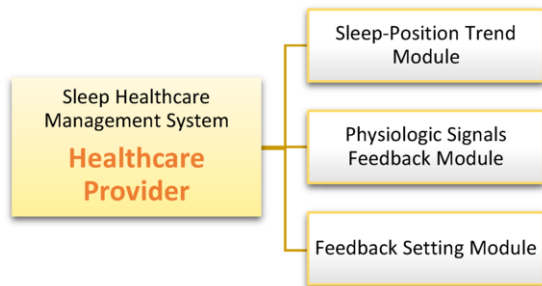


Figure 4 – System Structure: Sleep Healthcare Management System-Healthcare Provider.



Figure 5 – Physiological Chart: Mobile Application.

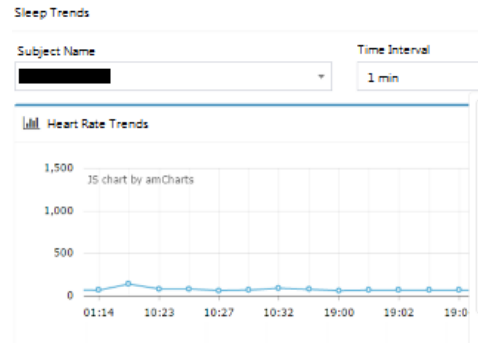


Figure 6 – Physiological Chart: Web Application.

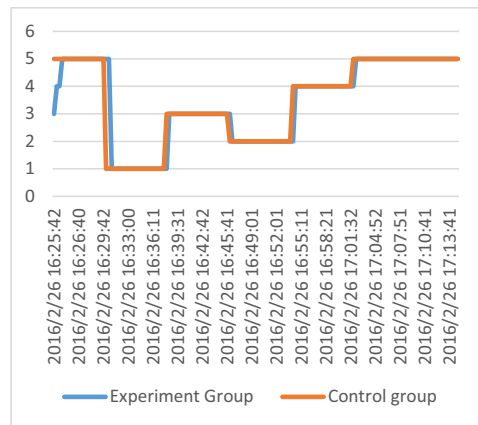


Figure 7 – Results of the Sleep-Position Detection: Sleep-Position Changes Trend.

Evaluation of the Sleep Healthcare Management System

This controlled trial was conducted in Taipei City from February 23, 2016 to March 1, 2016. We recruited 9 subjects who were attending university (20–35 years old; 2 male; 7 female).

All results are shown in Table 3. In general, data from the experiment and control group were consistent (Figure 7). However, there were problems noted with the results for the second, third, and fourth subject.

As for the second subject and the third subject, their accuracy of sleep detection is 0.81 and 0.74 respectively. The accuracy is also lower than other subjects because the breath tracker they wore, received signals with latency. If we are regardless of the phenomenon of signals with latency, the recorded state of both the second subject and the third subject is correct.

The accuracy of the sleep detection of the fourth subject is 73.51, much lower than other subjects. Because of the breath tracker she wore was interfered by noise at the beginning of the test. But the breath tracker will no longer be interfered by noise after the second state of the test.

The mean accuracy of our sleep-position detection algorithm is 0.89, the mean coefficient of correlation is 0.97. The result of this controlled trial showed that our system is highly reliable.

Table 3 – Accuracy and Coefficient of Correlation of Sleep-Position Detection.

Subject	Accuracy(%)	Coefficient of Correlation
1-Male	94.73	0.99
2-Female	81.29	0.97
3-Female	74.28	0.94
4-Female	73.51	0.90
5-Female	97.08	0.98
6-Female	93.91	0.98
7-Female	97.72	0.99
8-Female	91.42	0.99
9-Male	97.26	0.99
Mean	89.02	0.97

Discussion

The limitation of this study is the small size of the pilot evaluation. One of the next steps of this study is to do a randomized controlled trial for evaluating this “Sleep Healthcare System”. In this trial, the researchers would recruit more OSA patients and use a randomized program to assign these patients equally into the control group who are not using PSG and an experimental (intervention) group that are using the Sleep Healthcare System. Then the researchers would design a scale for the healthcare provider to know how efficiently and effectively this system could help them to monitor OSA patients.

Patients’ awareness of information security has increased and more attention is now paid to the privacy of personal data. How to ensure the security of data transmission and increase the reliability of system is an important question; therefore, the improvement of the security of the system would be an important future project.

Conclusions

Base on the system design, this study implemented the “Sleep Healthcare System,” which is suitable for obstructive sleep apnea patients.

This system has many features such as high convenience, real time, low cost; it also has intuitive and easy-to-use interface for adjusting the settings. It allows patients to do sleep monitoring at their home and enables health providers to get the sleep health condition of patients in a fast way.

People only need one smartphone and two wearable devices; then they can complete real-time sleep monitoring at home and they can view their physiological signals in a visualized way. Thus, people could know their sleep health condition quickly.

When the patient is experiencing an abnormal sleep health condition, the remote server can send feedback immediately and the healthcare providers can be made aware of the patient’s sleep health condition, enabling them to provide care services to the patients in a timely manner. In this context, this system would be useful for people who suffer from obstructive sleep apnea and their healthcare providers.

Acknowledgements

This project was mainly funded by the 9Rise International Mobile Health Technology Co., LTD., and partly supported by the national MOST project (MOST 104-2218-E-010-008-MY3).

References

- [1] K.A. Franklin and E. Lindberg, Obstructive sleep apnea is a common disorder in the population—a review on the epidemiology of sleep apnea, *Journal of Thoracic Disease* 7(8) (2015), 1311-1322.
- [2] Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine. Clinical Guideline for the Evaluation, Management and Long-term Care of Obstructive Sleep Apnea in Adults, *Journal of Clinical Sleep Medicine: JCSM: Official Publication of the American Academy of Sleep Medicine* 5(3) (2009), 263-276.
- [3] M.S. Aloia, J. T. Arnedt, L. Smith, J. Skrekas, M. Stanchina, and R.P. Millman, Examining the construct of depression in obstructive sleep apnea syndrome, *Sleep Medicine* 6(2) (2005), 115-121.
- [4] A.T. Mulgrew, N. Fox, N.T. Ayas, and C.F. Ryan, Diagnosis and initial management of obstructive sleep apnea without polysomnography: A randomized validation study, *Annals of Internal Medicine* 146(3) (2007) 157-166.
- [5] R.D. Chervin, D.L. Murman, B.A. Malow, and V. Totten, Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy, *Annals of Internal Medicine* 130(6) (1999), 496-505.
- [6] R. Ferber, R. Millman, M. Coppola, J. Fleetham, C. F. Murray, C. Iber, V. McCall, G. Nino-Murcia, M. Pressman, M. Sanders, et al., Portable recording in the assessment of obstructive sleep apnea: ASDA standards of practice, *Sleep* 17 (1994), 378-392.
- [7] J.M. Kortelainen, M.O. Mendez, A.M. Bianchi, M. Matteucci, and S. Cerutti, Sleep Staging Based on Signals Acquired Through Bed Sensor, *IEEE Transactions on Information Technology in Biomedicine* 14 (2010), 776-785.
- [8] W. Mei, T.B.J. Kuo, and C.C.H. Yang, Reliability Test of Our Self-Design Actigraphy on Autonomic Sleep/Wake Scoring in Human, National Yang-Ming University, Taipei, Taiwan, 2014.
- [9] J. Annuzzi Jr., L. Darcey, and S. Conder, Android Development Patterns: Best Practices for Professional Developer, Addison Wesley, 2015.
- [10] M. Cantelon, M. Harter, T.J. Holowaychuk, N. Rajlich, *Node.js in Action, 2nd Edition*, Manning Publications, 2013.

Address for correspondence

Polun Chang, PhD

Institute of Biomedical Informatics, National Yang-Ming University

Address: No. 155, Sec. 2, Linong St., Beitou District, Taipei City 112, Taiwan

E-mail: polunchang@gmail.com

Phone: 886-2-28267238

II. Health Data Science

This page intentionally left blank

Screening for Behavioral Risks: A Precision Healthcare Driven Approach for Chronic Pain Evaluation in Pediatric Specialty Care

Vibha Anand^a, Andrew S. Zeff^b, Steven J. Spalding^c

^aIBM T.J. Watson Research Center, Cambridge, MA, USA,

^bPediatric Institute, Cleveland Clinic, Cleveland, OH, USA ^cPhoenix Children's Hospital, AZ, USA

Abstract

We previously described an informatics tool (PROBE) to automate screening for behavioral risks for pain in children and adolescents. PROBE was deployed for a one year pilot study in our pediatric specialty care practice. Here we describe evaluation of this tool to assess self-report of pain, chronic disease activity and behavioral risks in 109 patients who sought routine care in the busy outpatient pediatric rheumatology practice of our large healthcare system. Results show that patients who self-report poorer self-efficacy and coping skills and night-time awakenings have significantly higher odds (8 and 5 times higher respectively) of reporting chronic pain even after accounting for their chronic disease activity. Our results show that automating screening in specialty care waiting rooms can not only inform the clinicians of patient's unknown risks but may even help drive the judicious use of precision healthcare resources such as cognitive behavioral therapy.

Keywords:

Chronic Pain; Pediatrics; Screening Behavioral Risks

Introduction

The medical science community has been studying pain for many decades including developing and evaluating pain questionnaires since the 1970s. [1–3] However, it is not until recently that focus on prevention, education and research has become a national priority in the US. [4;5] In the past, much work focused on addressing pain in life threatening settings such as cancer and palliative care. [6;7] However, as patient care became more collaborative, so did the use of technology to support management of pain in interdisciplinary care [8] and use of informatics-focused approaches to improve processes of care. [9] As the burden of chronic disease management (CDM) increases worldwide, there is a dire need to address pain in routine care. The field of pediatric medicine is not alone in this regard. Pain is the most distressing aspect of a chronic disease in children and can play a predominant role in their everyday lives.[10] Studies have shown that children who regularly experience pain perceive themselves to be more disabled, are more likely to restrict their activity and their pain related to disease commonly persists into adulthood. [11–13] For example, in a prospective observational study of children with Juvenile Idiopathic Arthritis (JIA), up to 76% of children reported pain on more than 60% of all days [11;14] and almost 40% of children documented pain every day. [14]

Despite the known long term effects of pain on quality of life, the assessment of pain experience in CDM varies. This is especially true in pediatric care [11;15] where studies have shown that assessment of pain experience in children and adolescents largely remains inconsistent. [16–18] A variety of factors in-

cluding busy clinical workflow or lack of evidence may contribute to a lack of complete and consistent pain experience assessment. However, it could also be that factors influencing overall pain experience may also be difficult to assess routinely, such as the behavioral risks for pain. Studies have shown that many behavioral risks may influence the overall pain experience in children and adolescents.[19–21] These include sleep deprivation, anxiety, depression, and painful conditions affecting a caregiver living in the child's home. [11;17;22] A child's pain experience is directly related to the number of family pain conditions, average level of parental pain and number of parental pain conditions [21]. Pain is exacerbated by changes in levels of stress, coping and mood [23] and differences in anxiety levels have shown to predict pain and participation levels in social and school activities [14]. Furthermore, when coping with higher levels of pain, children tend to use less positive self-statements and engage more in overly negative thinking about pain, i.e. catastrophizing statements. [24–26] Thus they are more likely to experience internalizing problems and have an overall poorer quality of life [27].

Fortunately, many evidence-based behavioral risk screening guidelines exist to implement complete and consistent assessment of pain routinely [28–30]. Therefore, if the many demands and constraints of the clinical workflow, including those with the use of modern electronic medical records (EMR) are met, these guidelines can be deployed in practice at the patient-level using a precision healthcare approach. For example, many commercial EMRs implement pain rating scales [31], however, very few if any support screening for behavioral risks in a CDM setting. Furthermore, lack of behavioral screening for pain is compounded because most health care systems lack the impetus and the resources to address it in the clinical workflow. Therefore, an informatics based complete and consistent pain assessment implementation outside the EMR may help in this regard.

We have shown in general pediatrics that evidence based screening can be efficiently implemented using an algorithmic approach in a computer-based clinical decision support system (CDSS). [32] Following this suite, we developed an automated screening solution, the Patient Risks Outcomes and Barriers Evaluation (PROBE) tool [33], to facilitate complete and consistent assessment of pain experience in young children and adolescents at their routine visit in a CDM setting. Although, external to our EMR, PROBE integrates patient's screening information in the clinical workflow, and gathers evidence for any further evaluation. In this pilot study, we 1) evaluate PROBE for screening behavioral risks for pain in a busy pediatric specialty care workflow; and 2) examine the relationship between self-report of chronic pain (CP), behavioral risks and clinically measured disease activity.

Methods

Patient Risks Outcomes and Barriers Evaluation (PROBE):

PROBE as a data capture tool has been described before. [33] It is implemented using the freely available Research Electronic Data Capture (REDCap) database platform [1;34], and commercially available electronic tablet devices (iPad). We added custom logic to deploy PROBE in pediatric rheumatology waiting rooms of our large healthcare system for screening patient's behavioral risks for CP.

Briefly, the iPad devices connect to the REDCap platform using a clinic-wide secure wireless connection. The web-browser interface on iPad provides for patient input to a screening form. The data corresponding to a form is captured and hosted on a dedicated server for protected health information (PHI). Customized logic ties it all together, for example calculates assessment scores on fields (where applicable) and links the patient screening, nurse and the provider worksheet forms. Please refer to [33] for details of these forms and associated clinical workflow. The latter captures clinical measurement(s) of patient's chronic disease activity from their EMR and is completed by a study coordinator (for aim 2). Clarifications on this form regarding patient's disease activity were sought from their attending physician if any. A summary of screening instruments implemented in PROBE is described below.

Behavioral Risks and Pain Rating Scales:

This study focuses on variables that are potentially modifiable for patients with existing clinical resources and treatment options. Therefore, we are using validated screening instruments as listed below. A single yes/no question was used for chronic pain assessment as listed below in (g).

- a. Childhood Depression Index and Screen for Child Anxiety Related Emotional Disorders (SCARED): [30]
- b. Bedtime problems, Excessive daytime sleepiness, Awakenings during the night, Regularity and duration of Sleep and Snoring (BEARS): [29]
- c. Parental Pain History Questionnaire (PPHQ): 12 item yes/no questionnaire directed towards the caregiver for parental pain conditions and if they received any treatment for it e.g. lower back pain, scored [28]
- d. Brief pain inventory scales: Numeric ratings to describe pain in the last week on a 0 (no pain) to 10 (pain as bad as you can imagine) scale. [35]
- e. Parent or patient reported global assessment (PaGA): "Considering all the ways that arthritis affects you / your child, please rate by selecting a number (in 0.5 increments) that best describes your / child's well-being: from 0 (overall well-being is very well) to 10 (overall well-being is very poor)." (*in consultation with pain experts at our institution*)
- f. Pain Coping Questionnaire: 3 items for measuring self-efficacy on a 5 point scale, scored. [28]
- g. Chronic Pain (CP) (self-report): "Have you / your child experienced pain more than 3 days per week for the last 3 months?" (yes/no) is asked of patient / parent to capture CP. (*in consultation with pain experts at our institution*)

Disease activity measure(s):

The American College of Rheumatology 30 (ACR30) criteria [36] measures patient's disease activity. To avoid pain as a confounder in ACR30, we use an alternative measure, the Juvenile Arthritis Disease Activity Score or JADAS for patient's disease

activity. JADAS-27 is a multifactorial measure and is responsive to clinically important differences. [37]

Dataset and Clinical Setting:

The PROBE tool was used in three practices of our tertiary care quaternary referral pediatric rheumatology practice. JIA is the most common chronic rheumatologic disease seen in the referral second opinion based practice. The data collected via PROBE are from April 2014 to April 2015 and the analyses are limited to patient visits where data exists for both screening and disease activity measurements during the study period.

Study outcomes of Interest:

Based on self-report of CP (yes or no), patient-visits (first visit for patients with multiple visits) were divided into two groups Study outcome(s) are association of a) *presence or absence of self-report of CP*; and b) *numeric parent-patient global self-assessment score (PaGA)*. Independent predictors are behavioral risks, pain rating scales and disease activity scores. This study was approved by our Institutional Review Board (IRB). All patient data were de-identified for statistical analyses.

Statistical Methods:

Patients with and without self-report of CP were compared on demographics and diagnosis using Wilcoxon rank sum, Chi-square, and Fisher's exact tests as appropriate. Visits at which patients reported CP were compared on clinical factors using univariable logistic regression models with generalized estimating equations (GEE) to adjust for the correlation between visits of the same patient. The associations between parent-patient global (self) assessment (PaGA) and pain coping scales were performed using the nonparametric Spearman correlation coefficients.

Models

We developed two prediction models from our dataset: a) *to predict self-report of chronic pain*; and b) *to predict PaGA (self-assessment) score*. To assess which demographic and clinical characteristics predict CP (at the visit level), multivariable logistic and linear regression models with GEE were created using backward selection on the following set of potential predictors: age, sex, five anxiety measures, three pain coping measures, five sleep problem measures, uveitis, normalized ESR, active joint count, and limited joint count. JADAS-27 score was used as a potential predictor of CP but not for PaGA (since PaGA is one of the components of the JADAS score). All analyses were performed on a complete-case basis. All tests were two-tailed and performed at a significance level of (p-value) 0.05 using SAS 9.4 software (SAS Inst., Cary, NC).

Results

A total of 148 visits from 109 patients were recorded during the study period. Patients generally visit in 3 month intervals (regular provider or infusion visits), 70% of patients had 1 visit record during the study period. Majority are female, with median age of 13 years. (Table 1)

Chronic Pain, Disease Activity and Pain rating scores:

More than half of the patients (n=56, 51%) self-report CP during their routine visits. (Table 1) For these patients, their disease activity score (JADAS-27) is 4 points higher (or worse) on average (Table 1); and their weekly pain rating scores are 3 points higher (or worse) (Table 2). When comparing CP to no CP group, there are significant differences in self-report of anxiety symptoms, sleep problems, average weekly pain scores, and pain coping skills, active joint count (AJC), physician global assessment (PGA), and parent-patient global assessment (PaGA) (Table 2). For subsequent analyses, we limit to patient's first visit in our record.

Chronic Pain (CP) Prediction Model

This model includes the following three variables – a pain coping variable (on a 1-5 scale with higher numbers indicating difficulty coping), sleep variable (yes or no), and a variable for Child Separation Anxiety Disorder (measured on the screened sub-scale). The odds of self-report of CP are 2 or more times higher (OR: 2.3, CI: 1.5 – 3.6, $p < 0.001$) for a unit change in pain coping scale score (to worse, i.e. to a higher score on a scale of 1 to 5). Thus, a 4 point difference in score (between the best and worst pain coping skill score) makes the odds for self-report of CP 8 or more times likely. Similarly, night time awakenings makes the odds 5 or more times likely (OR: 5.5, CI: 2.2 – 13.7, $p < 0.001$). In contrast a unit increase in measure of disease activity (JADAS-27) only increases the likelihood (of self-report) of CP by 30% (OR: 1.3, CI: 1.1 – 1.4, $p < 0.001$). The CP model also shows that patients with poorer pain coping skills have 60 – 90% probability of reporting CP when compared to 5 – 25% for patients with better coping skills even after adjusting for their disease activity (Figure 1)

Table 1 - Characteristics of the study population

Demographics	No Chronic Pain (N=53)	Chronic Pain (N=56)
Total Patients (N=109)	Median [Q1, Q3] or N (%)	Median [Q1, Q3] or N (%)
Age (years)*	13 [9,15]	14 [9,17]
Female	38 (72%)	43 (77%)
# of patient visits = 1	37(70%)	39 (70%)
Clinical Measures	No Chronic Pain (N=76)	Chronic Pain (N=72)
Total Visits (N=148)	Median [Q1, Q3] or N (%)	Median [Q1, Q3] or N (%)
Active Joint Count	0 [0,1]	1 [0,2]
Limited Joint Count*	0 [0,1]	0 [0,1]
Normalized ESR	4 (5%)	6 (8%)
Physician Global Assessment (PGA)*	0.0 [0.0,0.8]	1.0 [0.0,2.0]
JADAS-27 score*	1.0 [0.0,3.9]	5.0 [3.0,9.5]

*CP and No CP groups differ, p -value is significant

Patient/ Parent Global Assessment (PaGA) Prediction Model:

This model includes the following two variables – one pain coping variable (on a 1-5 scale with higher numbers indicating difficulty coping), and one sleep variable (yes or no). For a unit change to worse (higher) score for patient's pain coping skill level, the mean change is two units in PaGA score (adjusted mean effect: 2.0, CI: 1.3 – 3.1, $p = 0.003$). The spearman correlation coefficient between PaGA score and pain coping scale is 0.34 indicating low to moderate correlation. Self-report of patient problems such as falling asleep at bedtime increases the risk of self-report of CP to 3 or more times likely (adjusted mean effect: 3.6, CI: 1.5 – 8.5, $p = 0.003$). Pain coping skill scores predict measure of self-assessment and over all well-being. Please see Figure 2 details.

Table 2 - Self-report measures captured by PROBE

Factor	No Chronic Pain (N=76)	Chronic Pain (N=72)	p-value
Total Visits (N=148)			
Pain scores in last week and now: Median [Q1, Q3]			
WORST*	2 [0,4]	6 [5,7]	<0.001
LEAST*	0 [0,0]	2 [1,3]	<0.001

Factor	No Chronic Pain (N=76)	Chronic Pain (N=72)	p-value
Total Visits (N=148)			
AVERAGE*	1 [0,2]	4 [3,6]	<0.001
RIGHT NOW	0 [0,1]	3 [1,5]	<0.001
SCARED* score: N (%)			0.031
Total score 25 – 30	1 (1%)	5 (7%)	
Total score >30	2 (3%)	7 (10%)	
Child Generalized Anxiety Disorder Score = 9*	4 (5%)	13 (18%)	0.025
Child Separation Anxiety Disorder Score = 5*	3 (4%)	13 (18%)	0.012
Child Social Anxiety Disorder Score = 8*	5 (7%)	14 (19%)	0.027
Child Significant School Avoidance Score = 3*	1(1%)	14(19%)	<0.001
BEARS score: N (%)			
Overall Sleep Problems*	37 (51%)	57 (80%)	0.001
Problems falling asleep at bedtime*	13 (18%)	32 (44%)	0.001
Sleepy during the day*	24 (32%)	49 (69%)	<0.001
Wake up a lot at night or have trouble getting back to sleep*	11 (15%)	28 (39%)	0.001
Pain Coping Questionnaire: N (%)			<0.001
When you are in pain how often do you think you can do something to change it? (High=Never)*, N (%)			
1 - 2	40 (53%)	16 (22%)	
3 - 5	36 (47%)	58 (77%)	
How hard or easy is it for you to deal with being in pain? (High=Hard)*			
1 - 2	40 (53%)	18 (25%)	<0.001
3 - 5	36 (47%)	54 (75%)	
How often do you think you can do something to change your feelings when you are in pain? (High=Never)*			
1 - 2	40 (55%)	24 (34%)	0.012
3 - 5	36 (48%)	47 (66%)	
Parent-Patient Global Assessment: N (%)			
PaGA score, Median [Q1, Q3]*	0.5 [0.0,1.0]	3.5 [2.0,5.0]	<0.001

ξ Score of ≥ 25 may indicate presence of Anxiety Disorder. Scores higher >30 are more specific, *CP and No CP groups differ, p -value is significant

Discussion

Previous work has reported 30 - 40% of patients report chronic pain in a chronic condition like JIA. [22; 25] However, these findings are based on population-based sampling approaches, i.e. retrospective analyses of existing data or data outside the clinical visit (e.g. daily diaries). In contrast, our results from screening to implement a complete and consistent assessment of pain (experience) in pediatric specialty care shows that more than half of the patients (children and adolescents with a chronic condition) report chronic pain at routine visits. Additionally, we find that even after accounting for patient's clinically measured disease activity, report of poorer coping skills (self-assessment of pain coping skills) strongly predicts self-report of chronic pain and this association is even stronger with report of patient's sleep problems, e.g. night awakenings in adolescents (Figure 1). Furthermore, patient's pain coping skills

also largely predict their self-assessment score (for well-being), a well-recognized quality indicator of patient outcomes in CDM settings (Figure 2). Besides these symptoms of separation anxiety in younger children are other significant co-morbidities that associate with self-report of CP in our study. These findings have implications - not only for pain research in near future but also for addressing chronic pain, especially in pediatric subspecialties as also suggested by some other studies. Several studies have suggested that behavioral and cognitive therapies may be useful adjuncts in treatment of chronic conditions such as JIA. [25;38] Our findings corroborate with these previous studies and also add to it - a patient's coping skill level reflect their overall well-being and thus in turn sheds some light for individualized pain assessment and treatment strategy.

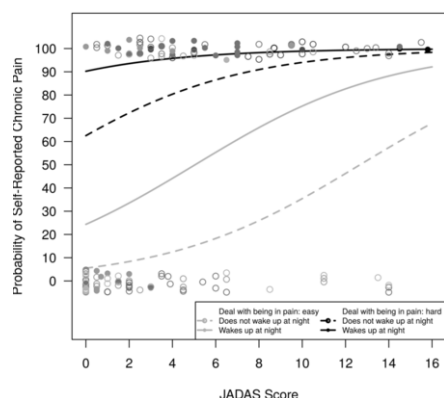


Figure 1 - Probability of self-report of chronic pain (CP). Lines represent predicted probability of self-reported chronic pain. Predictor variables: JADAS score, pain coping, and night awakening. Circles show observed patient-visit data (0 for no self-reported chronic pain, 100 for self-reported chronic pain). Open circles and dashed lines indicate patients with no night awakenings, and filled circles and solid lines indicate patients with night awakening problems. Shading of points and lines indicates self-reported coping on a 5 point scale from least difficult (light gray) to most difficult (black). Prediction lines are given only for the least and most difficult values.

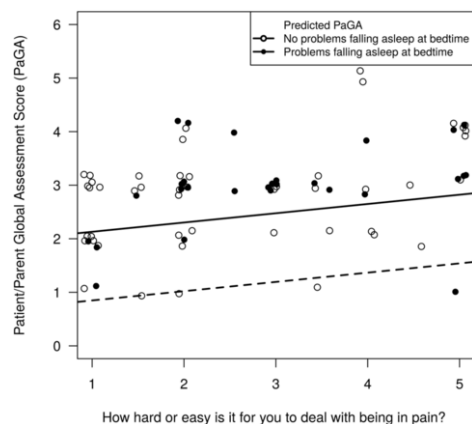


Figure 2 - Predicted score for Self-Assessment (PaGA). Lines represent predicted self-assessment score (PaGA). Predictor variables: pain coping, problems falling asleep at bedtime. Circles show observed patient-visit data. Open circles, dashed lines indicate patients with no problems. Filled circles, solid lines indicate patients with problems falling asleep.

As with all studies, our study also has some strengths and limitations. Behavioral risk assessments implemented in PROBE are applicable to a larger pediatric population. Therefore our findings may be different (perhaps stronger) for another chronic condition and speaks for the generalizability of the approach and the need for implementing care based quality indicators in routine practice. [39] Our limitations include a cross-sectional data analyses of self-report of CP linked to the clinical disease activity measures from patient's EMR for a given visit, e.g. physician global score and active joint counts. Therefore, it's possible that patient outcomes (chronic pain report, patient global score) may have improved in subsequent visits. However, we have not studied these outcomes longitudinally and the study is also limited by inherent selection bias - parent and patient perceptions of pain and pain experience may differ. [12] We have also not analyzed this dataset by the person reporting, i.e. parent vs. adolescent. Since the median age of patients in our study is 13 years, we infer that our results are more applicable to adolescent patients. Regardless, our approach makes a case for precision medicine for pain management in practice.

Pain management in precision healthcare driven approach:

Patient assessment scores and their interpretation can be presented to the busy clinicians in a visually meaningful and contextually enabling way - to drive patient-level care decisions such as to initiate adjunctive cognitive behavioral therapy or to refer for psycho-social or sleep medicine evaluation. However, both needs further research and perhaps an approach linking CDSS to patient's EMR. From a healthcare system's perspective, this approach may drive judicious allocation of clinical resources for optimal patient outcomes.

Conclusion

There is a dire need to implement screening for behavioral risks for chronic pain in children and adolescents in routine care, particularly in CDM settings. We presented an innovative approach to address individual's pain predictors in this study. Our findings suggest that enhancing a tool like PROBE into a functional CDSS to support risk identification and stratification for clinical decision making may be highly desirable for driving precision healthcare initiatives.

Acknowledgements

Dr. Anand received a grant from the Cleveland Clinic Foundation (CCF) to support this work. The authors wish to thank the following individuals: Susan and Art Karas for their valuable donations; and Sarah Worley and Anne Tang at QHS for providing statistical support. We would also like to thank the staff at the Center for Pediatric Rheumatology at Cleveland clinic for their help and support for this study.

Ethics Approval: This study was performed with approval from Cleveland Clinic Institutional Review Board (IRB).

References

- [1] B.M. Mount, R. Melzack, and K.J. MacKinnon, The management of intractable pain in patients with advanced malignant disease, *J. Urol.* **120** (1978) 720–725.
- [2] A.E. Reading, B.S. Everitt, and C.M. Sledmere, The McGill Pain Questionnaire: a replication of its construction, *Br. J. Clin. Psychol.* **21** (Pt 4) (1982) 339–349.
- [3] M.L. Zalon, Comparison of pain measures in surgical patients, *J. Nurs. Meas.* **7** (1999) 135–152.
- [4] (U.S.) Institute of Medicine, Committee on Comparative Effectiveness Research Prioritization. Initial National Priorities for Comparative Effectiveness Research. Washington, DC., *Natl. Acad. Press.* (2009).

- [5] Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research, *Mil. Med.* **181** (2016) 397–399. doi:10.7205/MILMED-D-16-00012.
- [6] N.D. Hughes, S.J. Closs, K. Flemming, and M.I. Bennett, Supporting self-management of pain by patients with advanced cancer: views of palliative care professionals, *Support. Care Cancer Off. J. Multinatl. Assoc. Support. Care Cancer.* (2016). doi:10.1007/s00520-016-3372-2.
- [7] K.E. Bickel, K. McNiff, M.K. Buss, A. Kamal, D. Lupu, A.P. Abernethy, M.S. Broder, C.L. Shapiro, A.K. Acheson, J. Malin, T. Evans, and M.K. Krzyzanowska, Defining High-Quality Palliative Care in Oncology Practice: An American Society of Clinical Oncology/American Academy of Hospice and Palliative Medicine Guidance Statement, *J. Oncol. Pract. Am. Soc. Clin. Oncol.* (2016). doi:10.1200/JOP.2016.010686.
- [8] C.E. Kuziemy, J.H. Weber-Jahnke, F. Lau, and G.M. Downing, An interdisciplinary computer-based information tool for palliative severe pain management, *J. Am. Med. Inform. Assoc. JAMIA.* **15** (2008) 374–382. doi:10.1197/jamia.M2519.
- [9] S.C. Ahluwalia, R.L. Leos, J.R. Goebel, S.M. Asch, and K.A. Lorenz, Provider approaches to palliative dyspnea assessment: implications for informatics-based clinical tools, *Am. J. Hosp. Palliat. Care.* **30** (2013) 231–238. doi:10.1177/1049909112448922.
- [10] T. Nghiem, J. Loui, S.C. Treherne, C.E. Anderson, A. Tsimicalis, C. Lalloo, J. Stinson, and K. Thorstad, Pain Experiences of Children and Adolescents with Osteogenesis Imperfecta: An Integrative Review, *Clin. J. Pain.* (2016). doi:10.1097/AJP.0000000000000395.
- [11] L.E. Schanberg, K.M. Gil, K.K. Anthony, E. Yow, and J. Rochon, Pain, stiffness, and fatigue in juvenile polyarticular arthritis: contemporaneous stressful events and mood as predictors, *Arthritis Rheum.* **52** (2005) 1196–1204. doi:10.1002/art.20952.
- [12] M.G. Sawyer, J.A. Carbone, J.N. Whitham, D.M. Robertson, J.E. Taplin, J.W. Varni, and P.A. Baghurst, The relationship between health-related quality of life, pain, and coping strategies in juvenile arthritis—a one year prospective study, *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **14** (2005) 1585–1598.
- [13] E. Limenis, H.A. Grosbren, and B.M. Feldman, The relationship between physical activity levels and pain in children with juvenile idiopathic arthritis, *J. Rheumatol.* **41** (2014) 345–51. doi:10.3899/jrheum.130734.
- [14] L.E. Schanberg, K.K. Anthony, K.M. Gil, and E.C. Maurin, Daily pain and symptoms in children with polyarticular arthritis, *Arthritis Rheum.* **48** (2003) 1390–1397. doi:10.1002/art.10986.
- [15] K.T. April, D.E. Feldman, R.W. Platt, and C.M. Duffy, Comparison between Children with Juvenile Idiopathic Arthritis (JIA) and their parents concerning perceived Quality of Life, *Qual. Life Res. Int. J. Qual. Life Asp. Treat. Care Rehabil.* **15** (2006) 655–661. doi:10.1007/s1136-005-3690-1.
- [16] W. Armbrust, J.G. Kaak, J. Bouma, O.T. Lelieveld, N.M. Wulfraat, P.J. Sauer, and E. van Sonderen, Assessment of disease activity by patients with juvenile idiopathic arthritis and the parents compared to the assessment by pediatric rheumatologists, *Pediatr. Rheumatol. Online J.* **11** (2013) 48. doi:10.1186/1546-0096-11-48.
- [17] M.H. Bromberg, K.M. Gil, and L.E. Schanberg, Daily sleep quality and mood as predictors of pain in children with juvenile polyarticular arthritis, *Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc.* (2011). doi:10.1037/a0025075.
- [18] L.C. Reigada, C.J. Hoogendoorn, L.C. Walsh, J. Lai, E. Szigethy, B.H. Cohen, R. Bao, K. Isola, and K.J. Benkov, Cross-Sectional Analysis of Anxiety Symptoms and Disease Severity in Children and Adolescents with Crohn's Disease, *J. Pediatr. Gastroenterol. Nutr.* (2014). doi:10.1097/MPG.0000000000000552.
- [19] T.M. Palermo, C. Eccleston, A.S. Lewandowski, A.C. Williams, and S. Morley, Randomized controlled trials of psychological therapies for management of chronic pain in children and adolescents: an updated meta-analytic review, *Pain.* **148** (2010) 387–97. doi:10.1016/j.pain.2009.10.004.
- [20] T.M. Palermo, and R. Kiska, Subjective sleep disturbances in adolescents with chronic pain: relationship to daily functioning and quality of life, *J. Pain Off. J. Am. Pain Soc.* **6** (2005) 201–207. doi:10.1016/j.jpain.2004.12.005.
- [21] L.E. Schanberg, K.K. Anthony, K.M. Gil, J.C. Lefebvre, D.W. Kredich, and L.M. Macharoni, Family pain history predicts child health status in children with chronic rheumatic disease, *Pediatrics.* **108** (2001) E47.
- [22] B. Margetic, B. Aukst-Margetic, E. Bilic, M. Jelusic, and L. Tambic Bukovac, Depression, anxiety and pain in children with juvenile idiopathic arthritis (JIA), *Eur. Psychiatry J. Assoc. Eur. Psychiatr.* **20** (2005) 274–276. doi:10.1016/j.eurpsy.2004.12.014.
- [23] J.N. Stinson, N.J. Luca, and L.A. Jibb, Assessment and management of pain in juvenile idiopathic arthritis, *Pain Res Manag.* **17** (2012) 391–6.
- [24] G.J. Reid, C.A. Gilbert, and P.J. McGrath, The Pain Coping Questionnaire: preliminary validation, *Pain.* **76** (1998) 83–96.
- [25] L.E. Schanberg, J.C. Lefebvre, F.J. Keefe, D.W. Kredich, and K.M. Gil, Pain coping and the pain experience in children with juvenile chronic arthritis, *Pain.* **73** (1997) 181–189.
- [26] M. Thastum, T. Herlin, and R. Zachariae, Relationship of pain-coping strategies and pain-specific beliefs to pain experience in children with juvenile idiopathic arthritis, *Arthritis Rheum.* **53** (2005) 178–84. doi:10.1002/art.21081.
- [27] N. Garnefski, H. Koopman, V. Kraaij, and R. ten Cate, Brief report: Cognitive emotion regulation strategies and psychological adjustment in adolescents with a chronic disease, *J. Adolesc.* **32** (2009) 449–54. doi:10.1016/j.adolescence.2008.01.003.
- [28] S. Kashikar-Zuck, A.M. Lynch, S. Slater, T.B. Graham, N.F. Swain, and R.B. Noll, Family factors, emotional functioning, and functional impairment in juvenile fibromyalgia syndrome, *Arthritis Rheum.* **59** (2008) 1392–1398. doi:10.1002/art.24099.
- [29] J.A. Owens, and V. Dalzell, Use of the “BEARS” sleep screening tool in a pediatric residents’ continuity clinic: a pilot study, *Sleep Med.* **6** (2005) 63–9. doi:10.1016/j.sleep.2004.07.015.
- [30] E. Taracki, I. Yeldan, E. Kaya Mutlu, S.N. Baydogan, and O. Kasapcopur, The relationship between physical activity level, anxiety, depression, and functional ability in children and adolescents with juvenile idiopathic arthritis, *Clin. Rheumatol.* (2011). doi:10.1007/s10067-011-1832-0.
- [31] C.P. Kaplan, C. Sison, and S.L. Platt, Does a pain scale improve pain assessment in the pediatric emergency department?, *Pediatr. Emerg. Care.* **24** (2008) 605–608. doi:10.1097/PEC.0b013e3181850c1c.
- [32] V. Anand, A.E. Carroll, P.G. Biondich, T.M. Dugan, and S.M. Downs, Pediatric decision support using adapted Arden Syntax, *Artif. Intell. Med.* (2015). doi:10.1016/j.artmed.2015.09.006.
- [33] V. Anand, and S.J. Spalding, Leveraging Electronic Tablets and a Readily Available Data Capture Platform to Assess Chronic Pain in Children: The PROBE system, *Stud. Health Technol. Inform.* **216** (2015) 554–558.
- [34] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J.G. Conde, Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support, *J. Biomed. Inform.* **42** (2009) 377–381. doi:10.1016/j.jbi.2008.08.010.
- [35] G. Tan, M.P. Jensen, J.I. Thornby, and B.F. Shanti, Validation of the Brief Pain Inventory for chronic nonmalignant pain, *J. Pain.* **5** (2004) 133–7. doi:10.1016/j.jpain.2003.12.005.
- [36] [1990 classification criteria of fibromyalgia from the American College of Rheumatology. Report of the Multicenter Criteria Committee], *Unión Médica Can.* **119** (1990) 272.
- [37] M. Bulatovic Calasan, L.D. de Vries, S.J. Vastert, M.W. Heijstek, and N.M. Wulfraat, Interpretation of the Juvenile Arthritis Disease Activity Score: responsiveness, clinically important differences and levels of disease activity in prospective cohorts of patients with juvenile idiopathic arthritis, *Rheumatol. Oxf.* **53** (2014) 307–12. doi:10.1093/rheumatology/ket310.
- [38] C.A. White, Cognitive behavioral principles in managing chronic disease, *West. J. Med.* **175** (2001) 338–342.
- [39] D.J. Lovell, M.H. Passo, T. Beukelman, S.L. Bowyer, B.S. Gottlieb, M. Henrickson, N.T. Ilowite, Y. Kimura, E.M. DeWitt, J. Segerman, L.D. Stein, J. Taylor, R.K. Vehe, and E.H. Giannini, Measuring process of arthritis care: a proposed set of quality measures for the process of care in juvenile idiopathic arthritis, *Arthritis Care Res Hoboken.* **63** (2011) 10–6. doi:10.1002/acr.20348.

Address for correspondence

Vibha Anand, PhD.
IBM T.J. Watson Research Center
75, Binney Street, Cambridge, MA 02142
anand@us.ibm.com

Is There a Priority Shift in Mental Health Clinical Trials?

Vibha Anand^a, Soumya Ghosh^a, Amit Anand^b

^a IBM T.J. Watson Research Center, Cambridge, MA, USA,

^b Center for Behavioral Health, Cleveland Clinic, Cleveland, OH, USA

Abstract

Mental health is the largest cost driver of healthcare. ClinicalTrials.gov is a web-based registry of trials conducted in human participants worldwide and serves as an important resource for both practice and research. We use the registry and dynamic topic modeling (DTM) methods to discover hidden topics and their evolution over last 17 years in trials in bipolar disorders. Our model suggests a “priority shift” from drugs to device and among populations studied in mental health trials. For example, transcranial magnetic stimulation for treatment resistant depression (TMS for TRD) trials have increased ($\chi^2 = 52.99$, $p < 0.001$, $n = 187$) while standardized drug safety and efficacy trials have remained constant ($\chi^2 = 0.72$, $p = 0.39$, $n = 101$) and more trials focused on pediatric and adolescents ($\chi^2 = 2.98$, $p = 0.08$, $n = 133$) may have been conducted since FDAAA 2007. We also derive unique data-driven insights in the discovered topic areas in this model based study.

Keywords:

Mood Disorders; Patient Selection; Natural Language Processing

Introduction

Mental health is an important societal issue. About 20% of adults and children are affected by some form of mental disorders in the US [1;2]. Globally, the costs for mental health care are greater than the costs for diabetes, respiratory disorders, and cancer combined and was estimated at \$2.5 trillion in 2010 [3]. These costs are expected to rise to \$6 trillion by 2030 [1;2]. Mental health clinical trials in the past studied safety and efficacy of therapeutic or behavioral interventions in many conditions ranging from schizophrenia to depression. However, it is not clear if there are emerging trends and topics in mental health trials since ClinicalTrials.gov registry was established.

The ClinicalTrials.gov (www.clinicaltrials.gov or registry) is a web-based registry and results database for clinical studies of human participants that was established in 2000. The registry and site are maintained by the National Library of Medicine (NLM) in the US but serves as an important resource for patient families, health care professionals and researchers globally. The FDA Amendments Act of 2007 (FDAAA) made trial registration and reporting of results mandatory and voluntary registration prior to this period enriched the registry with authoritative information on trials such as study design, intervention(s) and their outcomes in worldwide populations. A trial sponsor or its principal investigator typically registers the trial at the start and its record may be updated as needed. As of March 2017, 238,989 trials were registered of which only 24,799 (~10%) reported any results on the website

(www.clinicaltrials.gov). Given this valuable resource and that only about 10% of the studies have posted results, we need to understand the “overarching” themes of these clinical trials.

More recently, tools have been developed to navigate the registry site. However, most of the pre-baked tools provide summary information, and none provide a thematic view of the registry database. For example, it might be important to know if past mental health trials addressed conditions like schizophrenia or others such as depression and mania as well and in which populations. It may also be important to know whether these past trials have seen any “priority shift” in terms of treatment options and modalities. Such information may influence current and future guidelines for practice and spur new research directions [4-5]. In fact the Cochrane Collaboration, a volunteer-based organization leverages a volunteer workforce of 37,000 people for publishing systematic reviews for such purposes [6].

By analyzing word co-occurrence patterns, topic models can discover underlying themes or “topics” from collections of documents [7]. They use statistical algorithms to discover underlying themes, for example “political”, “sports” and “current affairs” from news articles. In this study, using dynamic topic modeling (DTM) methods [8], we analyze trial records related to “mania” and “bipolar disorders” in the registry - to discover topics from the last 17 years. Furthermore, we analyze topic trends over this time and trial characteristics such as sponsors, and participants associated with them.

Methods

A search using the keywords “Mania OR Bipolar Disorder” was initiated from the registry site on Nov. 24, 2016. The search resulted in 1220 trial records ranging from 1999 to 2016 and these were downloaded to a CSV file for further analyses. The details of variables comprising a trial record in the registry are given in Table 1. We used the “Title” and “First Received” variable fields in each record as inputs to learn a statistical model (please see Dynamic Topic Model section below) of word co-occurrence patterns that evolve over time. From this learned model, useful insights such as likely sponsors of a study area (topic), interventions studied and participant characteristics are derived, by analyzing individual documents (trial records) that meet certain threshold criteria.

Dynamic Topic Model

Topic models are generative latent variable models. They assume that the observed data is produced by a generative process, which is governed by hidden random variables. Thus the generative process defines a joint probability distribution over the observed and hidden random variables. The aim of the model is to learn conditional probability distributions or

posterior distributions over the hidden variables given the observed variables and the joint (probability) distribution.

Table 1 - Information in registry records (n=1220)

	Variables	Description
1.	NCT number	Unique trial record
2.	Title	Unique title record of trial
3.	Conditions	Disease condition(s) studied
4.	First Received	Date in years
5.	Age	Participant ages
6.	Gender	Both, Female, Male
7.	Recruitment	Status
8.	Interventions	Intervention studied
9.	Study Types	Interventional, Observational
10.	Study Designs	Randomized / Not Randomized, Safety / Efficacy etc.
11.	Study Results	Available / Not available
12.	Sponsor/Collaborators	Sponsor or Collaborator
13.	Phases	1-4
14.	Enrollment	Number of participants
15.	Funded By	Industry or Other
16.	Outcome Measures	Study outcome measure
17.	Acronym	Acronym for the study

A classic example of a topic model is Latent Dirichlet Allocation (LDA). LDA formally defines a “topic” as a distribution over a fixed vocabulary of words in a document collection. The observed variables in the model are the words in each document. LDA assumes that a document corpus expresses a set of K topics, where K is empirically chosen. Words in a document are then generated as such: a word in a document is drawn from one of the K topics, where the selected topic (topic assignment for the word) is chosen according to a per document distribution over topics. The goal then is to automatically discover the hidden structure of the model, i.e. “topics”, per document topic distribution, and per document per word topic assignment in the collection. In a LDA model each document exhibits all K topics, but in different proportions.

The inference problem is to learn the hidden structure given the observed words in the collection. LDA makes certain assumptions, first, the number of topics to be discovered are known, second, the ordering of words (“bag of words”) in the documents or for that matter ordering of documents in the collection does not matter. However, for large corpuses that run over years, such as from the registry, this assumption is not appropriate. Modeling the temporal order of documents in these collections is important for understanding topic evolution over time. Therefore, one approach is to use dynamic topic modeling methods (DTM) [8]. In DTM methods a “topic” is defined as a *sequence of distributions* over words rather than a single distribution of words. Thus, if a collection is arranged by years, each year is a “time slice” in DTM and documents with in each slice form a component topic model (e.g. LDA). This change facilitates tracking of thematic change in a collection over time or topic evolution and approximate structured variational inference techniques are used to infer an approximation to the posterior [9].

Model Selection and Introspection

We apply DTM to the “Title” and “First Received” fields in the downloaded dataset. Typically to learn a probabilistic model from the collection or “corpus”, a desired number of discoverable topics (empirically chosen or derived using other statistical techniques) along with initial parameters are given to the algorithm. We empirically chose to discover 8 topics from a corpus of 1220 trial records spanning over 17 years. We used an open source Natural Language Tool Kit (NLTK, www.nltk.org), and free python library for topic modeling (<http://pydoc.net/Python/gensim/0.11.1/gensim.models.wrapper.s.dtmmodel/>). NLTK was used to build the dictionary and corpus from the document collection and Gensim wrapper to set model parameters for learning the DTM. The learned model gives probability mass in document-topic matrix from which we calculated a) mean topic proportion per year or topic evolution over time (in years); and b) number of documents per topic by year. We used a χ^2 test to evaluate if the aggregate of mean topic proportions for a given topic before and after 2007 (FDA mandatory trial registration) has changed. By further analyzing individual documents (trial records) that meet a high threshold for a “given” topic (e.g. 0.9 in the document-topic matrix) we derived other useful insights. The threshold was chosen empirically and the results are described under the heading “Insights” in Results section. Among trial records expressing topics with the high-threshold probability we analyzed records for gender, recruitment status and study phase (based on 80th percentile cutoff).

Results

Our final model discovered 8 topics from the “Title” field in the registry records which were divided amongst 17 time slices based on the “First Received” study variable.

Topic Proportions and Evolution

Figure 1 describes the topic proportion and evolution over 17 years of registered trials. The topics are labeled T0 to T7 with human interpretation in Table 2 (by physician author – AA).

Mean topic proportion for every time slice is plotted in Figure 1a. As can be seen the document topic proportion varies over time which describes its evolutionary path. Such a variation would be expected in most topics due to changes in research focus, funding priorities or policy changes. Figure 1b describes the number of trials by topic in each year. As can be seen from here a large proportion of mental health trials were registered in 2005. This may correspond to the international committee of medical journal editors (ICJME) guidance of 2004, i.e. to require trial registration in consideration for publication. However, for some topics these variations are more pronounced, please refer to Table 3 for statistical results.

For example, topic T7 (transcranial magnetic stimulation for treatment resistant depression or TMS for TRD) trials form a larger proportion of mania and bipolar disorder trials since 2007 (Figures 1a and b). Since 2007, severe pediatric depression risk studies (T0) may have gained focus perhaps reflecting a health priority and the topic of “placebo controlled trials” in two drug standard treatment options lithium vs. lamotrigine (T3) for depression, has been relatively constant.

Insights

The following insights are derived from trial records exhibiting topic with high-threshold probability. Bolded headings indicate the variable field used for these analyses.

Study Types: The registry consists mostly of “interventional” trials, but there are some “observational” trials such as pediatric depression risk, acute psychosis, alcohol dependence, and cognitive therapy in children and adults (T0, T2, T4, T5, and T6 respectively).

Gender: Most mental health trials are conducted in both female and male genders, however, among Quetiapine (brand name Seroquel) treatment effect trials (T4), 20% are in males. Similarly, among trials studying cognitive therapy for weight (T6), 15% are in females only. (Figure 2)

Study Phase: Of note here, transcranial magnetic stimulation trials for treatment resistant depression are being conducted in all study phases. Cognitive behavioral therapy trials (T6) are in Phase 1, exception being trials in children and adolescents (T5), some of which are also in Phase 1/2 or Phase 4.

Recruitment Status: All placebo controlled drug trials for lithium vs. lamotrigine (T3) have been completed. About 85% of cognitive trials in adults (T6), and 90% of Olanzapine trials (T1) and acute psychosis treatment and symptoms trials (T2) have been completed. On the other hand, only about 50% of trials in children for severe risk of depression (T1) or cognitive therapy for anxiety (T5) have been completed. Trials of TMS for TRD are still recruiting and more than half of them have been completed.

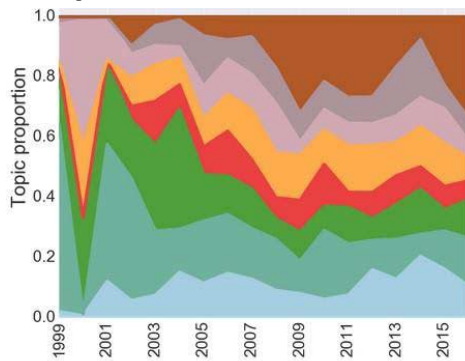


Figure 1(a) - Topic Proportion by year

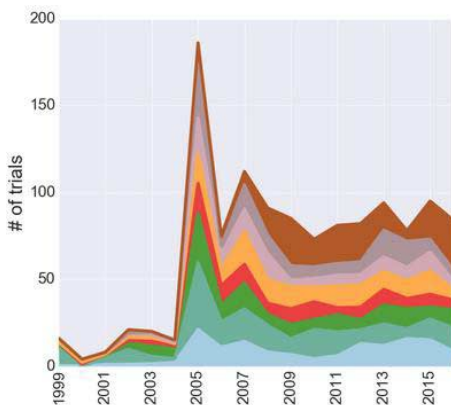
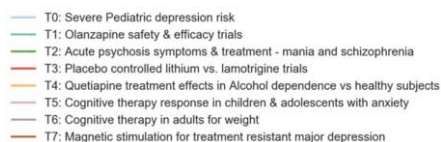


Figure 1(b) - # Trials per topic by year



Study Designs: Most trials are “Randomized” interventions. Some are “Observational” and these mostly study safety and efficacy (T0, T1, T2, T3). Other design parameters include case or cohort-control design (T1, T4, T6), group (T5), parallel crossover, double blind, open and non-open label.

Table 2 - Human interpretation on top words from model

Topic #	Top words in Topic	Clinical or Human Interpretation / Description
T0	'depressive', 'risk', 'pediatric', 'severe', 'youth', 'cognitive'	Severe Pediatric depression risk & cognition
T1	'safety', 'treatment', 'efficacy', 'manic', 'olanzapine', 'mania'	Safety & Efficacy studies for mania & manic episodes
T2	'mania', 'schizophrenia', 'symptoms', 'treatment', 'psychosis', 'acute'	Acute psychosis – mania, schizophrenia, symptoms & treatment
T3	'trial', 'versus', 'lithium', 'placebo', 'controlled', 'lamotrigine'	Placebo controlled and lithium and lamotrigine trials
T4	'quetiapine', 'treatment', 'effects', 'alcohol', 'dependence', 'healthy'	Alcohol dependence, Treatment using quetiapine. Studies with health subjects
T5	'adolescents', 'children', 'response', 'anxiety', 'clinical', 'cognitive'	Cognitive therapy response in children and adolescents with anxiety
T6	'therapy', 'adults', 'antipsychotic', 'based', 'weight', 'cognitive'	Cognitive Therapy in adults. Antipsychotics effect on weight.
T7	'depression', 'treatment', 'stimulation', 'magnetic', 'resistant', 'major'	Transcranial Magnetic Stimulation (TMS) for treatment resistant major depression (TRD)

Table 3 - Before - after 2007 trial comparison (by numbers)

Topic #	# of trials (n)	Chi-square (χ^2)	p-value
TMS for TRD (T7)	187	52.99	<0.001*
Severe pediatric depression risk (T0)	133	2.98	0.08*
Placebo controlled trials (T3)	101	0.72	0.39

* Significant at $p=0.05$

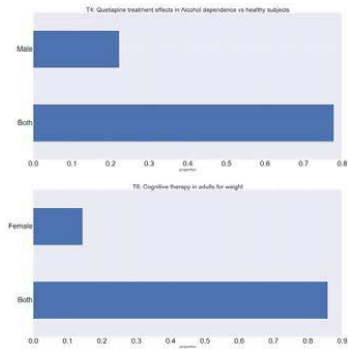


Figure 2 - Gender distribution in topics T4 and T6

Study Sponsor or Collaborators: Please refer to Figure 3.



Figure 3 - Sponsor/Collaborator of trials in "topics" (size of words reflect the frequency, Left-Right: T0 to T7)

Eli Lilly, the pharmaceutical and Massachusetts General Hospital (MGH) stands out in pediatric depression risk trials (T0) sponsored by the National Institute of Mental Health (NIMH). AstraZeneca, Merck and other pharmaceuticals stand out in Olanzapine safety and efficacy trials (T1) and Stanley Medical Research Institute for acute psychosis related trials (T2) sponsored by NIMH. Roxane and GlaxoSmithKline dominates placebo controlled lithium vs. lamotrigine trials (T3). NIMH also sponsors many of the alcohol dependence trials (T4), cognitive therapy trials (T5, T6) at MGH, Dartmouth Hitchcock, VA Palo Alto in the US. A larger proportion of TMS trials for TRD (T7) are done in "Research Universities" sponsored by NIMH, besides these there are trials in Denmark, Norway as well.

Experimental Interventions: Please refer to Figure 4. Compound LY2216684 and Ketamine may be studied in severe pediatric depression risk trials (T0) along with family focused cognitive therapy and mindfulness. Aripiprazole, Lithium, and Depakote may be studied in safety and efficacy

trials (T1). Tamoxifen is being studied in acute psychosis trials (T2). Behavioral interventions, and drugs Topiramate, Aripiprazole may be studied in depression trials (T3).

Compound AZD6765 and Naltrexone may be studied in alcohol dependence trials (T4). Ziprasidone along with Risperidone and Behavioral therapy may be studied in cognitive therapy trials for anxiety in children and adolescents. (T5) Mindfulness, Psychoeducation along with Cognitive therapy may also be studied in adults (T6) and Transcranial Magnetic Stimulation (TMS) is being studied for treatment resistant depression (T7) and most drug trials are placebo controlled.



Figure 4 - Interventions studied in topics (size of words reflects the frequency in each topic, Left-Right: T0 to T7)

Discussion

We discovered thematic areas or topics in mental health trials using DTM. Analyses of these topics provide unique insights into the hidden characteristics of the registry corpus. However, our study has some limitations. First, we selected 8 topics to discover empirically. It is possible that with a different number of topics, the model may produce finer or coarser thematic areas. An alternative approach to address the issue is to use a purely data-driven approach for number of topics such as the hierarchical Dirichlet process [9]. However, the chosen model is a good start for further improvement and the insights are from the "unknown". From the results, TRD using TMS trials started in larger numbers since 2007. This may be due to the commercial availability of TMS and its FDA approval in 2006. Trials of safety and efficacy with older drugs such as Standard treatment drug trials for depression show a downward trend [10-14]. Severe pediatric depression risk trials have gained focus in recent years [15-20] and cognitive therapy as an intervention is being studied even in the pediatric population perhaps reflecting a health priority. Additionally, newer compounds such as LY2216684 [21] and AZD6765 [22-24] are currently studied although they may be experimental in

nature. These findings can be used to drive resources for research in a precision healthcare system.

Conclusion

Insights from a text model based analyses of ClinicalTrials.gov registry reflects trend in the field of mental health trials, particularly in bipolar disorders. Our findings suggest that there may be a “priority shift” - from drugs to device based (TMS) intervention and to study adult as well as children and adolescent population in clinical trials in recent years. This may be due to several reasons for example, availability of new technologies, and due to high-level policies such as national health priorities and push towards precision healthcare driven systems.

References

- [1] NIMH » Any Disorder Among Children, (n.d.). <https://www.nimh.nih.gov/health/statistics/prevalence/any-disorder-among-children.shtml> (accessed December 15, 2016).
- [2] NIMH » Any Mental Illness (AMI) Among U.S. Adults, (n.d.). <https://www.nimh.nih.gov/health/statistics/prevalence/any-mental-illness-ami-among-us-adults.shtml> (accessed December 15, 2016).
- [3] T.R. Insel, Assessing the economic costs of serious mental illness, *Am. J. Psychiatry*. **165** (2008) 663–665.
- [4] C.L. Vale, L.H.M. Rydzewska, M.M. Rovers, J.R. Emberson, F. Gueyffier, and L.A. Stewart, Uptake of systematic reviews and meta-analyses based on individual participant data in clinical practice guidelines: descriptive study, *BMJ*. **350** (2015) h1088.
- [5] S. Gopalakrishnan, and P. Ganeshkumar, Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare, *J. Fam. Med. Prim. Care*. **2** (2013) 9–14.
- [6] About us | Cochrane, (n.d.). <http://www.cochrane.org/about-us> (accessed September 21, 2016).
- [7] D.M. Blei, Probabilistic Topic Models, *Commun ACM*. **55** (2012) 77–84.
- [8] D.M. Blei, and J.D. Lafferty, Dynamic Topic Models, in: Proc. 23rd Int. Conf. Mach. Learn., ACM, New York, NY, USA, 2006: pp. 113–120.
- [9] Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei, Hierarchical Dirichlet Processes, *J. Am. Stat. Assoc.* **101** (2006) 1566–1581.
- [10] M. Walpoth-Niederwanger, G. Kemmler, H. Grunze, U. Weiss, C. Hörtnagl, R. Strauss, I. Blasko, and A. Hausmann, Treatment patterns in inpatients with bipolar disorder at a psychiatric university hospital over a 9-year period: focus on mood stabilizers, *Int. Clin. Psychopharmacol.* **27** (2012) 256–266.
- [11] E. Vieta, J. Locklear, O. Günther, M. Ekman, C. Milteburger, M.L. Chatterton, M. Åström, and B. Paulsson, Treatment options for bipolar depression: a systematic review of randomized, controlled trials, *J. Clin. Psychopharmacol.* **30** (2010) 579–590.
- [12] M.L.M. van der Loos, P.G.H. Mulder, E.G.T.M. Hartong, M.B.J. Blom, A.C. Vergouwen, H.J.U.E.M. de Keyser, P.J.H. Notten, M.L. Luteijn, M.A. Timmermans, E. Vieta, W.A. Nolen, and LamLit Study Group, Efficacy and safety of lamotrigine as add-on treatment to lithium in bipolar depression: a multicenter, double-blind, placebo-controlled trial, *J. Clin. Psychiatry*. **70** (2009) 223–231.
- [13] S. Srivastava, and T.A. Ketter, Clinical relevance of treatments for acute bipolar disorder: balancing therapeutic and adverse effects, *Clin. Ther.* **33** (2011) B40–48.
- [14] M. Solmi, N. Veronese, L. Zaninotto, M.L.M. van der Loos, K. Gao, A. Schaffer, C. Reis, C. Normann, I.-G. Anghelescu, and C.U. Correll, Lamotrigine compared to placebo and other agents with antidepressant activity in patients with unipolar and bipolar depression: a comprehensive meta-analysis of efficacy and safety outcomes in short-term trials, *CNS Spectr.* **21** (2016) 403–418.
- [15] H.N. McKillop, and G.A. Banez, A Broad Consideration of Risk Factors in Pediatric Chronic Pain: Where to Go from Here?, *Child. Basel Switz.* **3** (2016).
- [16] W. Qubty, and A.A. Gelfand, Psychological and Behavioral Issues in the Management of Migraine in Children and Adolescents, *Curr. Pain Headache Rep.* **20** (2016) 69.
- [17] H. Denis, and A. Baghdadli, [Children and adolescents’ anxiety disorders], *Arch. Pediatr.* **24** (2017), 87–90.
- [18] A. Felhofer, O.D. Kothgassner, and C. Klier, How to Prevent Depression? Current Directions and Future Challenges in Children with Chronic Medical Conditions, *Psychiatr. Danub.* **28** (2016) 441–451.
- [19] R. Vameghi, S. Amir Ali Akbari, F. Sajedi, H. Sajjadi, and H. Alavi Majd, Path Analysis Association between Domestic Violence, Anxiety, Depression and Perceived Stress in Mothers and Children’s Development, *Iran. J. Child Neurol.* **10** (2016) 36–48.
- [20] M.A. Martin, S.K. Rothschild, E. Lynch, K.K. Christoffel, M.M. Pagan, J.L. Rodriguez, A. Barnes, K. Kara volos, A. Diaz, L.M. Hoffman, D. Plata, and S. Vilalando, Addressing asthma and obesity in children with community health workers: proof-of-concept intervention development, *BMC Pediatr.* **16** (2016) 198.
- [21] S. Dubé, M.A. Dellva, M. Jones, W. Kielbasa, R. Paudyal, A. Saha, and P. Rao, A study of the effects of LY2216684, a selective norepinephrine reuptake inhibitor, in the treatment of major depression, *J. Psychiatr. Res.* **44** (2010) 356–363.
- [22] G. Sanacora, M.R. Johnson, A. Khan, S.D. Atkinson, R.R. Riesenberger, J.P. Schronen, M.A. Burke, J.M. Zajecka, L. Barra, H.-L. Su, J.A. Posener, K.H. Bui, M.C. Quirk, T.M. Piser, S.J. Mathew, and S. Pathak, Adjunctive Lanicemine (AZD6765) in Patients with Major Depressive Disorder and History of Inadequate Response to Antidepressants: A Randomized, Placebo-Controlled Study, *Neuropsychopharmacology* **42** (2017), 844–853.
- [23] R. Machado-Vieira, I.D. Henter, and C.A. Zarate Jr, *New targets for rapid antidepressant action*, Progress in neurobiology (2015).
- [24] K.H. Bui, D. Zhou, F. Agbo, and J. Guo, Effect of multiple intravenous doses of lanicemine (AZD6765) on the pharmacokinetics of midazolam in healthy subjects, *J. Clin. Pharmacol.* **55** (2015) 1024–1030.

Address for correspondence

Vibha Anand, PhD,
IBM T.J. Watson Research Center
75, Binney Street, Cambridge, MA 02142
Email: anand@us.ibm.com

MedEx/J: A One-Scan Simple and Fast NLP Tool for Japanese Clinical Texts

Eiji Aramaki*, Ken Yano*, Shoko Wakamiya

Nara Institute of Science and Technology (NAIST), Japan

* Contributed equally to this work.

Abstract

Because of recent replacement of physical documents with electronic medical records (EMR), the importance of information processing in the medical field has increased. In light of this trend, we have been developing MedEx/J, which retrieves important Japanese language information from medical reports. MedEx/J executes two tasks simultaneously: (1) term extraction, and (2) positive and negative event classification. We designate this approach as a one-scan approach, providing simplicity of systems and reasonable accuracy. MedEx/J performance on the two tasks is described herein: (1) term extraction ($F_{\beta=1} = 0.87$) and (2) positive-negative classification ($F_{\beta=1} = 0.63$). This paper also presents discussion and explains remaining issues in the medical natural language processing field.

Keywords:

Medical Informatics; Natural Language Processing

Introduction

Medical reports using electronic media are now replacing those of paper media. Correspondingly, information processing techniques in the medical field have radically increased their importance. Especially, natural language processing of numerous medical reports and obtaining knowledge from them has drawn much attention for its promise of facilitating precise and timely treatments [1].

Our goal is to promote the development of practical tools that support medical decisions. For medical reports in English, several tools have become available, such as the MAYO Clinic Vocabulary Server, MedLEE [2], and cTAKES [3]. In contrast, tools for Japanese texts are not yet available. Considering that most clinical text is written in the respective domestic languages of users, it is reasonable to develop language-specific processing tools for each language.

To process Japanese texts, this paper presents a proposal of a Japanese clinical NLP tool, designated as **MedEx/J**, having the following two functions.

1. **Entity Recognition:** This task identifies expressions of diseases and complaints in clinical texts. This task is similar to a named entity recognition task. Herein we name this task **ER**.
2. **Positive-Negative (P/N) Classification:** Because clinical texts include various events, not only the current disease (**positive disease**) but also earlier diseases and suspected diseases, events do not always occur. We regard such an entity that does not occur in a target patient as a **negative entity**. We designate the task of classification of a positive entity and negative entity as **P/N Classification**.

These workflows are presented in Figure 1. The respective tasks are important for medical NLP. For example, an Automatic

Content Extraction (ACE) information extraction program deals with event extraction, by which each event is annotated with temporal and modal markers. A similar effort was made in Time relation (TimeML) [4] and the Penn Discourse TreeBank (PDTB) [5] projects (discourse relation). Usually, a state-of-the-art extraction method is based on Conditional Random Fields (CRFs) [6].

The P/N classification presents an important issue because negative events are frequently described in clinical texts. Negative events have two characteristics. First, various words (or sometimes phrases) affect the negation of events. Another is that a word that triggers a negation sometimes has a long scope. Because of the observation presented above, a heuristic approach suffers from its handling of various negative patterns, thereby motivating the use of machine learning methods for negative detection [7-9].

Compared to these earlier works, our method has two important differences that are readily apparent: (1) character-based processing, and (2) the fusion of two tasks (both entity recognition and P/N classification).

- **Character-based NLP:** Traditional NLP studies regard a word as a minimum unit. A clinical text, however, includes numerous complex nouns consisting of multiple Chinese characters. This complexity often leads to errors of morphological analyses. Therefore, this study uses a character-based method with no preprocesses.
- **Fusion task:** Usually, the P/N classification task is applied after the first entity recognition step. However, we combine these two tasks into one because we consider that the required information for P/N classification resembles that of the entity recognition.

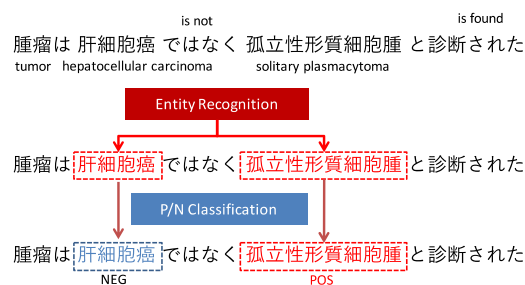


Figure 1 - Workflow of MedEx/J.

In addition to the benefits described above, these two features dramatically simplify the system. One system can be built with no extra systems. In compensation for the system simplicity, the

task must solve a complex task, term identification with P/N classification, which invariably produces large feature spaces. We designate this approach as a *one scan* approach because multiple tasks are solved under one sequential labeling.

The experimentally obtained results reveal that, in spite of the risk of large feature space caused by our one-shot approach, our method has achieved high performance, (1) term extraction ($F_{\beta=1} = 0.87$) and (2) positive negative classification ($F_{\beta=1} = 0.63$), demonstrating the feasibility of the proposed approach.

Materials: The NTCIR Clinical Text Dataset

To implement the proposed method, we use a corpus consisting of the NTCIR shared task data [10] and its compatible inhouse dataset. In these data, the symptom and diagnosis related expressions are marked with $\langle p \rangle$ $\langle /p \rangle$ (for a positive entity) or $\langle n \rangle$ $\langle /n \rangle$ (for a negative entity) as follows.

- **Compound noun:** noun compound word was marked as a whole.
- **Verbal phrase:** verbal phrases were **not** marked.
- **Disease identification test:** the disease was marked if and only if the existence of the certain virus represents a single particular disease.
- **General description:** when the expression describes ‘general’ information related to the disease and/or the name of the clinic, it was **not** marked.
- **Non-alphabetical character:** non-alphabetical characters and non-numerical characters (e.g. “↓”) were marked with the prior noun phrases if and only if the marks represented the conditions and were attached to the previous noun phrases to form the names of the disease.
- **Modality related word:** Words and phrases that suggest modalities of the symptoms (e.g. positive 陽性, negative 陰性, prevention 予防) were not included in tags. However, some modality-related words were included if and only if they were connected to compose the standard disease names.

For detailed information related to the annotation scheme, see the annotation guideline available on the web[13]. We used data consisting of 468 (9,286 sentences in total) Japanese documents.

Methods

Character-based labeling

Using the corpora described in the material section, we adjusted the proposed NLP tool. In our task scheme, an input is a sentence. The output is a sentence with $\langle p \rangle$ (for positive) or $\langle n \rangle$ (for negative). This output represents two tasks: (1) entity recognition and (2) P/N classification. The first task is represented as the tagged scope. The second task is represented as the type of tag ($\langle p \rangle$ or $\langle n \rangle$).

The ER is to mark the beginning (b), the inside (i), and the outside (O) labels over input text sequences. Differences between the proposed method (character-based) and the existing method (word-based) sequence labeling are explained below (Fig. 2).

(a) Input text

腫瘍は肝細胞癌ではなく肝の孤立性形質細胞腫と診断された。

It was diagnosed as solitary plasmacytoma, not hepatocellular carcinoma.

(b) Character-based sequence labeling

腫瘍は肝細胞癌ではなく孤立性形質細胞腫と診断された
O O O B-I-I-I O O O B-I-I-I-I-I-I-I-I O O O O O O

(c) Word-based sequence labeling

腫瘍は肝細胞癌ではなく孤立性形質細胞腫と診断された
O O B-N-I-N-I-N O O O B-P-I-P-I-P-I-P O O O O O O

(d) Output text with tags

腫瘍は<n>肝細胞癌</n>ではなく肝の<p>孤立性形質細胞腫</p>と診断された。

It was diagnosed as <p>solitary plasmacytoma</p>, not <n>hepatocellular carcinoma</n>.

Figure 2 - Sentence representation in sequential labeling: (a) input, (b) character-based, (c) word-based, and (d) output

Usually, a word-based approach uses information from a word unit, such as the part of speech and word inflection as features. Our approach, however, could not use them because it uses no morphological analysis. Instead, we used script-type information of each character. In Japanese, the script type usually indicates word boundaries.

Figure 3 specifies the template for generating feature functions used respectively for the character-based approach (Fig. 3(a)) and word-based approach (Fig. 3(b)). In the tables, ‘Char Position’ or ‘Word Position’ are the relative positions from the current character or word token. In the character-based CRF template, ‘Char’ and ‘Char type’ respectively stand for the surface character and character types: *kanji*, *hiragana*, *katakana*, and *alphanumeric*. The ‘N-gram’ in the column headers specifies which n-grams among uni-gram, bi-gram or tri-gram are used. The window sizes of the respective settings are chosen to achieve the best performance in preliminary experiments.

(a) Character-based features

Feature Type	N-gram	Window Size (unit : character)
Char	1, 2-gram	-2,-1,0,1,2,3,4,5
Char Type	1, 2-gram	-2,-1,0,1,2,3,4,5

(b) Word-based features

Feature Type	N-gram	Window Size (unit : word)
Word	1,2,3-gram	-2,-1,0,1,2
Word[-1]	1,2,3-gram	-2,-1,0,1,2
Word[-2:-1]	1,2,3-gram	-2,-1,0,1,2
Yomi	1,2-gram	-2,-1,0,1,2
POS	1,2-gram	-2,-1,0,1,2

For word-based CRF, features of five types were used. The feature types, ‘Word’, ‘Word[-1]’ and ‘Word[-2:-1]’ respectively specify the surface word, previous word, and previous two words. The ‘Yomi’ is the hiragana reading of the surface word; ‘POS’ is the part of speech of the corresponding word.

Figure 3 - Feature functions template (a) for character-based CRF and (b) for word-based CRF

Fusion Method: ER with P/N classification

The P/N classification is usually applied to the ER results. In contrast, we propose a one-shot approach to solving these two problems simultaneously by handling positive and negative events as independent negative entities.

Negative events are associated with negative or suspicious predicate phrases, emerged in generic or unrelated context with patients. For negative entities, we use ‘*B-N*’ and ‘*I-N*’ labels, respectively, instead of ‘*B-P*’ and ‘*I-P*’. Following are some samples of negative predicate phrases that frequently emerged in our corpus.

- <N>悪性腫瘍<N>が**疑われた**
 - (**suspicious of** <N>malignant tumor<N>)
- <N>発熱<N>は認め**なかった**
 - (<N>fever<N> **not found**)
- <N>肺結核 <N>は**否定的だった**
 - (<N>pulmonary tuberculosis<N> **was negative**)

It is noteworthy that only the negative entity is annotated. Negative trigger phrases, such as “*suspicious of*” and “*not found*”, are captured explicitly by machine learning.

Results

Corpus and Setting

For evaluation, we collected another set of 500 (10,266 sentences in total) annotated Japanese discharge summaries, which were not used to construct the model for CRF. We used the CRF toolkit with standard parameters[14]. We then investigated the feature contribution to the performance in learning settings of two types: word-based and character-based.

To compare the performance against a word-based method, we implemented both methods.

Evaluation metrics

Performance of the first task (ER) was assessed using the **precision**, **recall**, and **F-measure** ($\beta=1$) according to earlier studies. Precision is the percentage of correct entities found by a participant’s system. Recall is the percentage of entities present in the corpus that were found by the system. The *F*-measure is the harmonic mean of precision and recall. An entity is regarded as correct only if it was an exact match of the corresponding entity in the data file. The evaluation method is the same as that used for the CoNLL-2000 shared task. A Perl script used for evaluation is available at the CoNLL-2000 website[15].

We adopted evaluations of two types. ER (only) and ER+P/N classification (fusion task). ER was a complaint and diagnosis or not: only entity. Modality included modalities of two types: positive or negative.

Table 1 presents performance information of the ER task associated with character-based method and word-based method. The proposed character-based method outperformed word-based method in all three metrics. Table 2 presents performance of the fusion method: ER + P/N classification. In this task, the character-based method also outperformed the word-based method in all metrics. The N-tag performance was not good compared with that of P-tag, suggesting that P/N classification is a rather difficult task. This observation is reasonable because P/N classification sometimes becomes a difficult decision during annotation by humans.

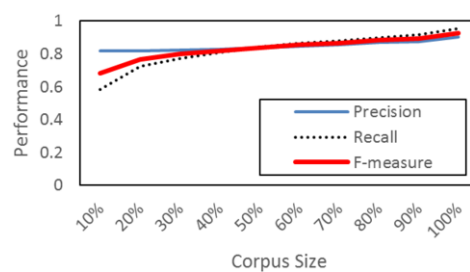
Table 1 - ER performance

	Precision	Recall	F-measure
Character-based method	0.911	0.826	0.866
Word-based method	0.920	0.799	0.855

Table 2 - ER + P/N classification performance

	P/N	Precision	Recall	F-measure
Character-based method	<P>-tag	0.853	0.815	0.833
	<N>-tag	0.735	0.447	0.554
Word-based method	<P>-tag	0.854	0.793	0.823
	<N>-tag	0.676	0.333	0.444

(a) <p>-tag



(b) <n>-tag

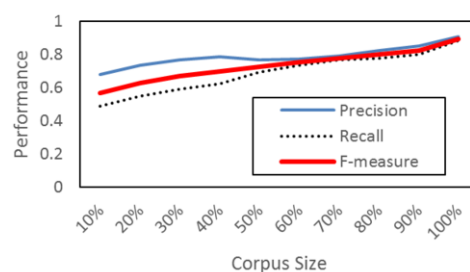


Figure 4 - Performance and the corpus size.

Discussion

Next, several remaining issues are investigated: the corpus size effect, entity coverage, and processing time.

Corpus size effect

It is important to ascertain how many data are required for training. To investigate this point, the performance of character-based methods for the fusion ER task was analyzed by changing the corpus size from 10% to 100% in 10% steps. Figure 4(a) presents the performance change of <p>-tag extraction. Figure 4(b) shows that of <n>-tag extraction. The performance of each increases gradually as the corpus size increases. The recall rate tends to be more affected by the corpus size than by the precision rate. From these figures, we inferred that the performance did not reach a plateau even when using the entire corpus, therefore suggesting that some room exists for improvement by increasing the corpus size.

Entity coverage: Instance level analysis

Figure 5 presents the type or token coverage of the top n frequent terms. The figure shows a typical long-tail shape in which the top 500 terms occupy only 50% of tokens, which implies that many varieties of terms appear in the clinical texts, also motivating the corpus increase.

Processing Time

Performance and processing speed are important indicators for practical systems. In our estimation, a large hospital such as a university hospital handles 3,000 patients daily, generating 60,000 sentences per day. We checked the CPU time of a processor (core i7 6800K 6core/12thread 3.4 GHz (Turbo Boost 3.6 GHz)) and its associated memory (32GB (8GB×4) DDR4-2133 Quad-Channel). The time to process 1,000 sentences was 2 min 44 s for training and 2 s for tests, which indicates that 120 (=2×60,000/1000) s is sufficient for all patients in one day. We regard this as a practically feasible time for most hospitals.

Future task

An important problem persists: mapping the extracted term to the concept, such as ICD-10 or SNOMED-CT [12]. We are also trying to develop this module incorporating the orthographic disambiguation techniques. Another remaining task is to use a smarter method such as Recurrent Neural Network (RNN). However, we think our one-scan approach presents certain benefits in terms of system simplicity.

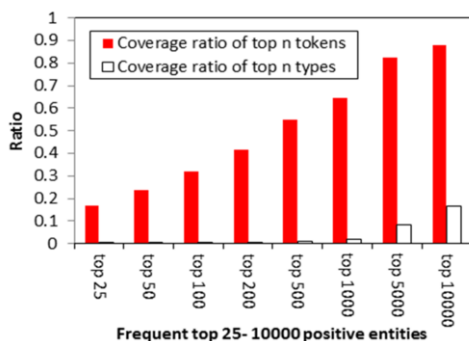


Figure 5 - Coverage ratio of top 25–10,000 frequent token (red) or types (white) of entities. Coverage ratio is calculated as follows: Coverage ratio = (accumulated frequency of top n types or tokens) / (total frequency of all types or token in the training corpus)

In summary, room for improvement remains, requiring more novel approaches, error analyses, and more effort for this task to meet its needs.

Conclusions

We developed a Japanese clinical tool, **MedEx/J**, consisting of two functions: (1) term extraction and (2) P/N classification. Our approach executes both tasks as one sequential labeling. This approach has two important benefits: 1) Preprocessing was simplified because morphological analysis was not required. 2) Its performance was much better (term extraction ($F_{\beta=1} = 0.87$) and positive negative classification ($F_{\beta=1} = 0.63$)) than that of a word-based method although using a much simpler feature set for the CRF.

Our results suggest that the approach is extremely effective for handling medical reports. Additionally, we believe that this simple implementation promises great benefits from a practical

viewpoint, such as convenience of installation into hospitals, maintenance costs, and the development of commodity applications.

The software is available at <http://sociocom.jp/parser.html>.

Acknowledgements

This study was supported (in part) by JSPS KAKENHI Grant Numbers JP16H06395 and 16H06399, and Health and Labour Sciences Research Grant No. 28030301.

References

- [1] W.W. Chapman, P.M. Nadkarni, L. Hirschman, L.W. D'Avolio, G.K. Savova, and O. Uzuner, Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions, *J Am Med Inform Assoc.* 18 (2011) 540–543.
- [2] C. Friedman, S.B. Johnson, B. Forman, and J. Starren, Architectural requirements for a multipurpose natural language processor in the clinical environment, *Proc Annu Symp Comput Appl Med Care* (1995), 347–351.
- [3] cTAKES, <https://ctakes.apache.org/>
- [4] J. Pustejovsky, J. Castaño, R. Ingria, R. Sauri, R. Gaizauskas, A. Setzer, and G. Katz, Timeml: Robust specification of event and temporal expressions in text, *Fifth International Workshop on Computational Semantics (IWCS-5)* (2003).
- [5] R. Prasad, N. Dinesh, A. Lee, A. Joshi, and B. Webber, Annotating Attribution in the Penn Discourse Tree-Bank, *Proceedings of the International Conference on Computational Linguistics and the Annual Conference of the Association for Computational Linguistics (COLING/ACL2006) Workshop on Sentiment and Subjectivity in Text*, 2006, pp. 31–38.
- [6] J. Lafferty, A. McCallum, and F. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, *Proceedings of the International Conference on Machine Learning (ICML2001)*, 2001, pp. 282–289.
- [7] Y. Huang and H.J. Lowe, A novel hybrid approach to automated negation detection in clinical radiology reports, *J Am Med Inform Assoc* 14 (2007) 304–311.
- [8] P.G. Mutalik, A. Deshpande, and P.M. Nadkarni, Use of general-purpose negation detection to augment concept indexing of medical documents: A quantitative study using the UMLS, *Journal of the American Medical Informatics Association* 8 (2001) 598–609.
- [9] W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan, Evaluation of negation phrases in narrative clinical reports, *Proc AMIA Symp*, 2001, pp. 105–109.
- [10] M. Morita, Y. Kano, T. Ohkuma, M. M., and E. Aramaki, Overview of the NTCIR-10 MedNLP task, *Proceedings of NTCIR-10*, 2013.
- [11] SNOMED-CT. 2002. SNOMED Clinical Terms Guide. College of American Pathologists.
- [12] Eiji Aramaki, Takeshi Imai, Kengo Miyo, and Kazuhiko Ohe, Orthographic Disambiguation Incorporating Transliterated Probability, *International Joint Conference on Natural Language Processing (IJCNLP2008)*, 2008, pp. 48–55.
- [13] NTCIR MEDNLP shared task: <http://mednlp.jp/ntcir11/>
- [14] CRF++: <https://taku910.github.io/crfpp/>
- [15] CoNLL2000: <http://www.cnts.ua.ac.be/conll2000/chunking/>

Address for Correspondence

Eiji ARAMAKI, Ph.D. <aramaki@is.naist.jp>

Locally Adaptive Operators for Red Lesions Detection in Eye Fundus Images

Luiza Dri Bagesteiro^a, Daniel Welfer^b, Marcos Cordeiro d'Ornellas^b,
Juliano F. Kazienko^c, Carlos Jesus Pereira Haygert^d, Gustavo Nogara Dotto^d

^a Universidade Federal do Pampa, RS, Brasil,

^b Department of Applied Computing, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil,

^c Colégio Técnico Industrial de Santa Maria, Universidade Federal de Santa Maria, Santa Maria, RS, Brasil,

^d Department of Clinical Medicine - Santa Maria University Hospital - Federal University of Santa Maria (UFSM), Rio Grande do Sul, Brasil.

Abstract

One of the major features required by automated software tools of screening for diabetic retinopathy is the detection of red lesions. This paper presents a new automatic method in order to locate red lesions in color eye fundus images. The method relies on mathematical morphology operators and has a coarse and a fine detection stages, respectively. The former detection stage detects structures of low-intensity values in the retina, such as microaneurysms, hemorrhages, blood vessels and the fovea center. Additionally, the latter stage proposes to improve the detection of red lesions identified in the previous stage. For experiments, we use the well-known publicly available DIARETDB1 database. The results indicate that our method detected red lesions with 75.81% and 93.48% of mean sensitivity and mean specificity, respectively.

Keywords:

Diabetic Retinopathy; Retina; Microaneurysm

Introduction

Evaluation of the retina can provide information regarding the presence of ocular diseases such as glaucoma and macular degeneration. Moreover, chronic systemic diseases such as high blood pressure and diabetes can also be detected through of the retina conditions [1]. Furthermore, diabetes contributes to decrease the health-related quality of life among patients [2]. One consequence of diabetes is vision impairment, or even total vision loss (i.e., blindness). Diabetic Retinopathy (DR) consist in a complication caused by diabetes, and may damage the visual acuity of people at working age [3]. The clinical signs that characterize the Diabetic Retinopathy are the presence of microaneurysms and hemorrhages (i.e., red lesions) that are observable by digital eye fundus photography. Thus, the early detection of red lesions can prevent or delay visual acuity loss in patients with Diabetic Retinopathy.

The main retinal features and red lesions (e.g. microaneurysms and hemorrhages) are illustrated in Figure 1 for a typical retinal image. However, the automatic detection of red lesions is still difficult for several reasons, namely:

1. the red lesions have different sizes and shapes [1];

2. it is difficult to distinguish red lesions from other retinal structures such as blood vessels and the fovea region [4, 5];
3. there are slow background variations and illumination artifacts in the fundus image which contributes to decreases the accuracy of detection of red lesions. Red lesions are quite difficult to detect manually as well, even for an expert in ophthalmology [6, 7].

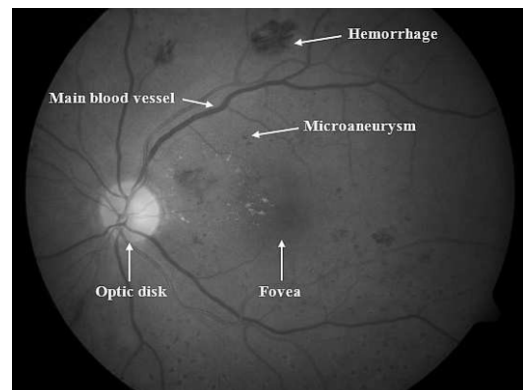


Figure 1 – Retinal main physiological structures and lesions in a typical digital eye fundus photo.

The purpose of this paper is to present a novel method for automated detection of red lesions in eye fundus images. In order to demonstrate our results and to avoid subjective visual evaluation, we used quantitative measurements to assess the accuracy of our red lesions detection method.

Methods

The images of the DIARETDB1 database (Standard Diabetic Retinopathy Database Calibration level 1, version 1 [8], are used to validate experimentally our proposed red lesions detection method. The DIARETDB1 database is available on the WEB, and consists in a total of 89 color eye fundus images of 1500×1152 pixels at 24bit RGB. Moreover, we use reference images (i.e., the ground truth) to evaluate the detection results obtained by our method. It is important to remark that the DIARETDB1 has 53 images

(i.e., 59.55% of images) containing red lesions (e.g., microaneurysms and hemorrhages).

First, an adaptive contrast enhancement technique is applied to the green channel f_g of the original DIARETDB1 color eye fundus image (e.g., Figure 2(a)). In this stage, we apply the Contrast Limited Adaptive Histogram Equalization (CLAHE) transform to achieve our enhancement results [9, 10]. In this way, a new image f_1 is obtained as detailed in Equation 1:

$$f_1 = \text{CLAHE}(f_g), \quad (1)$$

In our experiments, we have used 10x10 blocks with a normalized clip limit of 0.2 for CLAHE. The f_1 image is illustrated in Figure 2(b).

Now, on the f_1 image, we apply the method proposed by Walter *et al.* [11] to remove slow background variations, and to improve the contrast. This technique uses alternating sequential filters based on morphological closings, ϕ , and morphological openings, γ , so as to estimate the background f_2 of the image f_1 , as shown in Equation 2. The morphological opening, $\gamma^{(B)} = \delta^{(B)}[\varepsilon^{(B)}(f_1)]$, of the image f_1 is the consequence of the erosion ε of f_1 by a defined structuring element B, followed by the dilation δ of the result of the erosion ε of f_1 by the same structuring element. Conversely, $\phi^{(B)} = \varepsilon^{(B)}[\delta^{(B)}(f_1)]$ is the morphological closing of the image f_1 (i.e., consequence of a dilation, δ , followed by an erosion, ε , of f_1 by a defined structuring element B) [12].

$$f_2 = (((f_1 \phi B) \gamma B) \cdots \phi nB) \gamma nB, \quad (2)$$

where B is a flat disc-shaped structuring element with a fixed radius of 8 pixels, and n is the number of iterations ($n = 2$ in our experiments) on which the sequential combination of opening and closing is applied. As next step, the image difference between f_2 and f_1 is computed. Then, a new image f_3 without slow background variations is obtained, as detailed in Equation 3. Figure 2(c) depicts the image f_3 .

$$f_3 = f_2 - f_1, \quad (3)$$

Additionally, we can identify low intensity regions (e.g., dark areas that potentially are associated with red lesions) using the grayscale reconstruction by dilation as shown in Equation 4

$$f_4 = R_{f_1}(\delta^{(B)}(f_1)), \quad (4)$$

where f_1 and $\varepsilon^{(B)}(f_1)$ are the mask and marker images, respectively. In addition a flat disc-shaped structuring element with a fixed radius of 10 pixels was used in the marker image. Figure 2(d) depicts the image f_4 obtained from the aforementioned described operator.

Afterwards, in order to flatten more the areas related to red lesions, we remove large intensity peaks by applying a grayscale reconstruction by erosion R^* , as shown in Equation 5. Then, taking $\delta^{(B)}(f_4)$ as a marker image, and f_1 as a mask image, the reconstructed f_5 image is obtained, as shown in Figure 2(e).

$$f_5 = R_{f_1}^*(\delta^{(B)}(f_4)), \quad (5)$$

Soon after, the image difference between f_5 and image f_1 is computed, as show below:

$$f_6 = f_5 - f_1, \quad (6)$$

The output f_6 is an image where structures like thin vessels, large vessels, hemorrhage, and microaneurysms are enhanced against the background of the color eye fundus images. Figure 2(f) depicts the image f_6 .

At this point, the difference between the H-minima and the H-maxima transforms of f_6 (i.e., $H\min_H(f_6)$ and $H\max_H(f_6)$, respectively) are computed according to Equation 7. The image resulting from the difference between $H\min_H(f_6)$ and $H\max_H(f_6)$, namely f_7 , is illustrated in Figure 2(g).

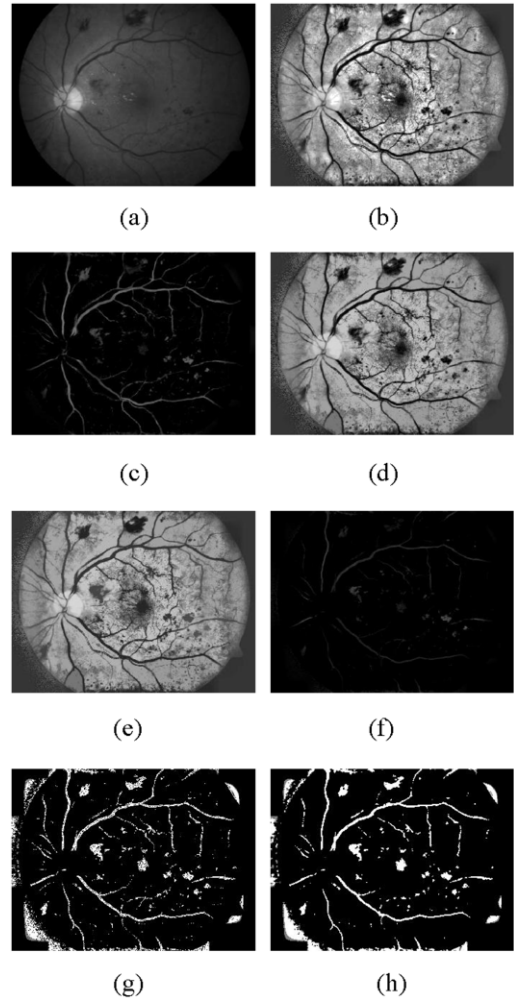


Figure 2 – Coarse detection stage. (a) Green channel image. (b) Enhanced image f_1 . (c) Image f_3 without slow background variations. (d) Reconstruction by dilation using f_1 as the mask image. (e) Reconstruction by erosion; (f) The difference between f_5 and f_6 image (image f_6). (g) Difference between the $H\min_H(f_6)$ and $H\max_H(f_6)$ followed by a grayscale inversion (f_7 image). (h) After a grayscale closing and opening operation (image f_8).

The H-minima transform removes all connected basins with grayscale heights (i.e., contrast) less than h . Thus, the output of the H-minima transform is an image where structures like thin or large vessels and hemorrhage are removed. Conversely, the H-maxima transform removes all connected peaks with contrast less than h , resulting in an H-maxima image usually containing only a few basin structures corresponding to peaks higher than h . Thus, subtracting the H-minima from the H-maxima increases the contrast of vessels and red lesions (that usually show low contrast). This processing stage will be important later so as to detect the red lesions without using thresholding techniques.

$$f_7 = (Hmin_h(f_6) - Hmax_h(f_6))^c, \quad (7)$$

where, H -minima and H -maxima transforms removes the connected basins and peaks. Afterwards, in order to flatten more the red lesions region, we remove large intensity basins and peaks by applying a grayscale closing, ϕ , and a grayscale opening, γ , respectively (see Equation 8). Figure 2(h) depicts the resulting image f_8 .

$$f_8 = \gamma^{(B)}[\phi^{(B)}(f_7)], \quad (8)$$

To improve the red lesions detection accuracy, we apply the toggle morphological contrast enhancement method to image f_8 [12]. As described by Equation 9, this method uses a grayscale erosion and a grayscale dilation operation.

$$f_9(x,y) = \begin{cases} \varepsilon^{(B)}(f_{18}), & \text{if } f_8(x,y) - \varepsilon^{(B)}(f_{18}) \geq \delta^{(B)}(f_{18}) - f_8(x,y) \\ \delta^{(B)}(f_{18}), & \text{if } f_8(x,y) - \varepsilon^{(B)}(f_{18}) < \delta^{(B)}(f_{18}) - f_8(x,y) \end{cases} \quad (9)$$

where, B is a diamond-shaped structuring element with a fixed radius of 3 pixels.

Finally, in order to detect the red lesions candidate regions we employed the RMAX operator (i.e., which detects the regional maxima pixels) [13]. The regional maxima, RMAX, returns a binary image from the f_9 grayscale image. If a set of connected pixels of f_9 , namely $S(f_9)$, have a constant intensity value bigger to its neighboring pixels, then the logical value "1" is assigned to such set of pixels. Otherwise it is assigned zero to this set of pixels. The RM AX image can be found according the Equation 10.

$$f_{10} = RMAX(f_9) = f_9 + 1 - R_{f_9+1}(f_9). \quad (10)$$

In order to identify and remove the red lesions regions which are actually regions of vessels, the circularity of the shape of the regions is calculated. The central idea is that the regions belonging to the network of vessels have an elongated shape, and therefore low circularity. Thus, among the n boundary shapes, we select that one with the highest compactness as described by Equation 11.

$$f_{11} = \begin{cases} \text{maintain region,} & \text{if } 4\pi \frac{A}{P^2} > \lambda \\ \text{remove region,} & \text{Otherwise,} \end{cases} \quad (11)$$

where P is the boundary shape perimeter, A the shape area and λ is the threshold used to remove undesirable elongated shapes.

Next, we describe the next stage used for improving the accuracy of the red lesions boundary detection, based on the foreground pixels of the binary image f_{11} . A marker

image f_{12} is obtained according to Equation 12, based on images f_{11} and on the resulting image from the union between image f_3 and f_6 . This marker image f_{11} is illustrated in Figure 3(b).

$$f_{12}(x,y) = \begin{cases} 0, & \text{if } \delta^{(B)}(f_{11}(x,y)) = 1 \\ f_3(x,y) \vee f_6(x,y), & \text{Otherwise,} \end{cases} \quad (12)$$

where B is a diamond-shaped structuring element. A dilation using this large structuring element guarantees that the red lesion region is included in f_{12} . Then, using f_{12} as a marker image and $f_3(x,y) \vee f_6(x,y)$ as a mask image, the reconstructed image f_{13} is obtained, as described in Equation 13 and illustrated in Figure 3(c). Hence, only regions outside of the red lesions are reconstructed.

$$f_{13} = R_{f_3(x,y) \vee f_6(x,y)}(f_{12}). \quad (13)$$

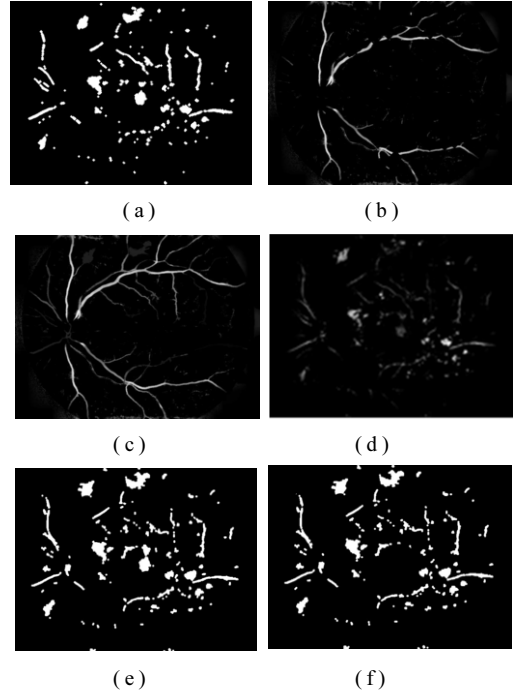


Figure 3 – Fine red lesions detection. (a) Result after the identification and removal of the vessel network (image f_{11}). (b) Marker image f_{12} . (c) Reconstructed image f_{13} . (d) f_{14} image. (e) Resulting image f_{18} containing the final red lesions detected. (f) Binary image f_{18} without the macula region in the foreground.

Then, the red lesions region is enhanced by subtracting the previously reconstructed image f_{13} from $f_3(x,y) \vee f_6(x,y)$ (see Eq. 14), and the result f_{14} is illustrated in Fig. 3(d).

$$f_{14} = [f_3(x,y) \vee f_6(x,y)] - f_{13}. \quad (14)$$

Finally, in order to detect the final red lesions we apply again the steps shown by equation 7, 8, 9 and 10 in that

order, on the image f_{14} . In other words, we obtain four new images namely: 1) f_{15} ; 2) f_{16} ; 3) f_{17} ; and 4) f_{18} (i.e., Fig. 3(e)). Afterwards, as post-processing stage we identified and removed the macula from image f_{18} [19]. The macula is the darker region of the image, and is located near to the optic disk center. Then, as the macula is a region of low intensity, it can be confused with the red lesions. Thus, the remotion of the fovea region provides more precision for the proposed method [14]. Figure 3 (f) depicts the resulting image f_{18} .

Figure 4 illustrates the methods used in coarse detection stage. Figure 5, illustrates de summarized steps used to improve the detection of red lesions.

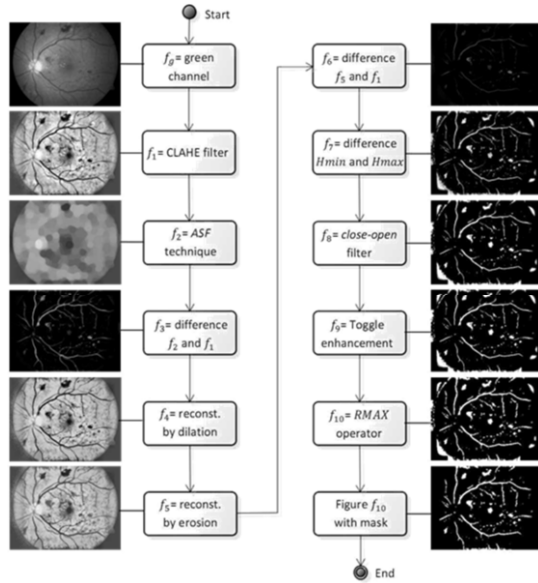


Figure 4– Summarized preliminary steps to detect red lesions.

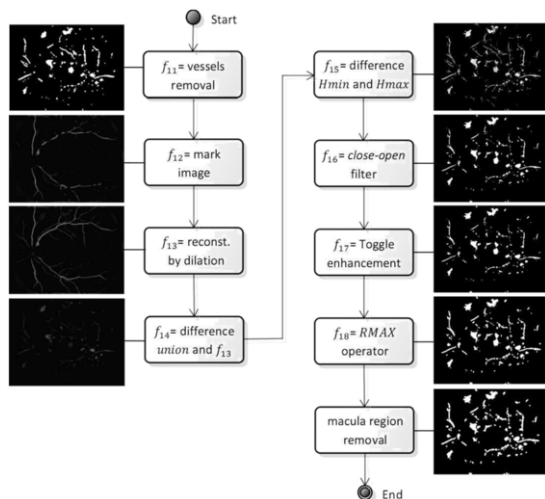


Figure 5 – Summarized steps used to improve the detection of red lesions.

Results and Discussion

We evaluate the performance of the proposed method using metrics as follows: the measurements of sensitivity, specificity, positive predictive value and misclassified proportion measures [15, 16] on 89 DIARETDB1 fundus images. The sensitivity is given by the proportion of true positives, TP, with respect to proportion of false negative, FN, pixels (i.e., $=TP/(TP+FN)$). The specificity is the proportion of true negative, TN, pixels with respect to the proportion of false positive, FP, pixels (i.e., $TN/(TN+FP)$). The predictive value is the proportion of true positives, TP, with respect to proportion of false positive, FP, pixels (i.e., $TP/(TP+FP)$). As indicated in Table 1, the proposed method achieved 75.81% of average sensitivity, and 93.48% of average specificity considering all 53 images with red lesions in the DIARETDB1 database. For the sake of comparison, Table 1 summarizes the red lesions detection results achieved by our method in the coarse and fine stage, indicating that in the fine stage the method significantly improves the results achieved in the coarse stage. It shall be observed in Table 1 that our method also was tested on retinal images not containing red lesions, achieving an average specificity of 94.65% and an average misclassified proportion of 5.35%.

Table 1 – Performance comparison between coarse and fine red lesions detection (DIARETDB1 database)

Detection Stage	Retinal Images with Red Lesions ^a							
	average sensitivity	average specificity	average predictive value	average misclassified proportion	average TP	average FN	average FP	average TN
Coarse	64.63%	93.23%	5.81%	6.72%	1423.34	716.1	20630.23	284430.4
Fine	75.81%	93.48%	6.6%	6.5%	1661.80	477.57	19842.59	285218.1
Detection Stage	Retinal Images without Red Lesions ^b							
	average sensitivity	average specificity	average predictive value	average misclassified proportion	average TP	average FN	average FP	average TN
Coarse		93.74%		6.26%	0	0	19215.61	287984.4
Fine		94.65%		5.35%	0	0	16430.81	290769.2

^a Comprises 53 images of the DIARETDB1 database.

^b Comprises 36 images of the DIARETDB1 database. The sensitivity and predictive value cannot be calculated because all TP (True Positives) and FN (False Negatives) values in color eye fundus images without red lesions are equal to zero.

The method proposed in this work is based on mathematical morphology and it presents important features. Initially, the method relies on adaptive morphological operators, as the operator of maximum regionals. It is applied to two stages of red lesion detection, namely, preliminary detection and refined detection, adapting according to the processed image. Hence, basic threshold is discarded in our method because it does not take into consideration image peculiarities. Another remark regarding the method is its robustness. The method presented in this work is not negatively affected by the presence of lesions or even from optical disc in detection phase, since it identifies red lesions discriminating regions with low and high intensity. Low intensity regions are related to potential red lesions. On the other, it is possible to find detection methods in literature (e.g., Badea *et al.* [4]) that carry out firstly the detection of all existing structures within image, and afterwards they remove the optical disc and white lesions. Such approach is more susceptible to imprecisions in detection process. It is important to point out the method presents limitations regarding blood vessels removal and identification of little lesions in image. We observed through our experiments that the detection and removal

of blood vessels affected both sensitivity and specificity measurements. The phase of blood vessels extraction contributed to sensitivity loss due to the frequent proximity between lesions and vessels. Thus, when the circularity is calculated it takes place the removal of the lesions because several lesions are jointed to blood vessels. This is a drawback compared to a method of blood vessels detection that identifies the skeleton tree of them based on their linear structure. Considering the specificity, the blood vessels removal was limited due to little vessels. Particularly, the method did not remove all these kind of vessels from the image. In the image, the pixels corresponding to such little vessels are identified as FP and they decrease the specificity in the proposed method.

Conclusion

We describe a method for detecting red lesions in retinal images. The proposed method comprises a stage which increases the detection accuracy by removing dark structures such as vessel and the macula region, and by means of a further refinement of the red lesions candidate regions. Our experimental results based on the DIARETDB1 public database indicate that our method achieved 75.81% of average sensitivity and 93.48% of average specificity. It shall be observed, that the great challenge in the case of detection of red lesions is to achieve a high average value of sensitivity and specificity. Our proposed method demonstrated this very clearly through a complete evaluation, i.e., without ignoring the main measures of accuracy described in the literature. In addition, it is only when a particular method achieves high sensitivity and high specificity that the good result (i.e., close to the result specified by the ground truth) is reached. Future work will focus on the development of other methods related to the automatic detection of diabetic retinopathy. Additionally, we intend to investigate regarding the use of red lesions along with another retina elements in people identification.

References

- [1] M. D. Abràmoff, M. K. Garvin, M. Sonka, Retinal imaging and image analysis, *IEEE Reviews In Biomedical Engineering* 3 (2010) 169–208.
- [2] N. Kaur, J. Kaur, M. Accharya and S. Gupta, Automated detection of red lesions in the presence of blood vessels in retinal fundus images using morphological operations, 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), *Delhi, 2016*, pp. 1–4.
- [3] T. A. Ciulla, A. G. Amador, B. Zinman, Diabetic retinopathy and diabetic macular edema, *Diabetes Care* 26 (2003) 2653–2664.
- [4] P. Badea, D. Danciu, L. Davidescu, Preliminary results on using an extension of gradient method for detection of red lesions on eye fundus photographs, in: *IEEE Int. Conf. Automation, Quality and Testing, Robotics*, volume 3, Clujnapoca, 2008, pp. 43–48.
- [5] N. Silberman, K. Ahlrich, R. Fergus, L. Subramanian, Case for automated detection of diabetic retinopathy, in: *Association for the Advancement of Artificial Intelligence (AAAI) Spring Symposium: Artificial Intelligence for Development*, AAAI, Palo Alto, 2010, pp. 85–90.
- [6] H. F. Jaafar, A. K. Nandi, W. Al-Nuaimy, Automated detection of red lesions from digital colour fundus photographs, in: *International Conference IEEE Engineering in Medicine and Biology Society*, Boston, 2011, pp. 6232–6235.
- [7] L. Seoud, T. Hurtut, J. Chelbi, F. Cheriet and J. M. P. Langlois, "Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening," in *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1116–1126.
- [8] T. Kauppi, V. Kalesnykiene, J.-K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, J. Pietilä, DIARETDB1 diabetic retinopathy database and evaluation protocol, in: *Medical Image Understanding and Analysis (MIUA)*, 2007, pp. 61–65.
- [9] G. Yadav, S. Maheshwari and A. Agarwal, Contrast limited adaptive histogram equalization based enhancement for real time video system, 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), New Delhi, 2014, pp. 2392–2397.
- [10] B. Kurt, V. V. Nabiyeu, K. Turhan, Medical images enhancement by using anisotropic filter and clahe, in: *International Symposium of Innovations in Intelligent Systems and Applications (INISTA)*, 2012, pp. 1–4.
- [11] T. Walter, J.-C. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy - detection of exudates in color fundus images of the human retina, *IEEE Transactions on Medical Imaging* 21 (2002) 1236–1243.
- [12] Shih, F. Y. *Image Processing and Mathematical Morphology: Fundamentals and Applications*. New York: CRC Press, 2009.
- [13] Landini, G., Randell, D.A., Fouad, S. and Galton, A. (2017). Automatic thresholding from the gradients of region boundaries. *Journal of Microscopy*.
- [14] D. Welfer, J. Scharcanski, D. R. Marinho, Fovea center detection based on the retina anatomy and mathematical morphology, *Computer Methods and Programs in Biomedicine* 104 (2011) 397 – 409.
- [15] Gacesa R, Barlow DJ, Long PF. Machine learning can differentiate venom toxins from other proteins having non-toxic physiological functions. *PeerJ Computer Science* (2016).
- [16] Sahithi, S., & G. P. (2016). Feature segmentation of blood vessels using active contour with hybrid region information with application to retinal images. *I-Manager's Journal on Image Processing*, 3(1), 37–43.

Address for correspondence

Corresponding author: Daniel Welfer – daniel.welfer@ufsm.br

Semi-Automatic Mark-Up and UMLS Annotation of Clinical Guidelines

Matthias Becker^a, Britta Böckmann^{a,b}

^aDepartment of Medical Informatics, University of Applied Sciences and Arts, Dortmund, Germany

^bIMIBE, University Hospital Essen, Germany

Abstract

Clinical guidelines and clinical pathways are accepted and proven instruments for quality assurance and process optimization in the healthcare domain. To derive clinical pathways from clinical guidelines, the imprecise, non-formalized abstract guidelines must be formalized. The transfer of evidence-based knowledge (clinical guidelines) to care processes (clinical pathways) is not straightforward due to different information contents and semantical constructs. A complex step within this formalization process is the mark-up step and annotation of the text passages to terminologies. The Unified Medical Language System (UMLS) provides a common reference terminology as well as the semantic link for combining the clinical pathways to patient-specific information. This paper proposes a semi-automated mark-up and UMLS annotation for clinical guidelines by using natural language processing techniques. The algorithm has been tested and evaluated using a German breast cancer guideline.

Keywords:

Decision Support Systems, Clinical; Natural Language Processing; Machine Learning

Introduction

Clinical guidelines identify, summarize and evaluate the highest quality of evidence and most current data about prevention, diagnosis, prognosis, and therapy, including the dosage of medications and cost-effectiveness [1]. They aim to help healthcare professionals and patients to make the best decisions about treatment and care for a particular condition or situation. The content of a guideline is based on a systematic review of the clinical evidence—the main source of evidence-based care. A decisive factor for the success and use of clinical guidelines is the provision of knowledge at the point of care. The main challenge can be summarized as follows: Imprecise, non-formalized abstract guidelines have to be implemented as concrete processes. The acceptance of clinical decision support systems is only given if they are integrated into the clinical workflow and presented at the point of care [2]. Clinical pathways, tools used to guide evidence-based healthcare, are appropriate for that purpose and have been implemented internationally.

Due to different information contents and semantical constructs, the transfer of evidence-based knowledge (clinical guidelines) to care processes (clinical pathways) is not straightforward. Although there are already several approaches to formalize clinical guidelines to create clinical pathways, the process of formalization and the annotation of terminologies are very complex [3].

For example, during the mark-up step, the extraction of all pertinent guideline recommendations, which should be

considered during pathway development, and classification (e.g. medication, procedure, observation, additional information) of this information are done manually by a health professional. Since the guidelines update at different intervals, this time-consuming step must be performed again after each update. The idea is to semi-automate the mark-up process by natural language processing techniques and simultaneously annotate the marked passages by UMLS [4, 5]. This terminology makes it possible to connect the components (generic knowledge and patient-specific information) with semantically rich links in a real-time dynamic manner at the point of care and with less effort [6]. It has been demonstrated that clinical guidelines used as decision support systems significantly improve patient care and are effective instruments to decrease undesired practice variability. However, acceptance of clinical decision support systems is only given if they are integrated into the clinical workflow and presented at the point of care [7]. Model-based approaches to support the development of guideline-compliant pathways, such as PathGuide, enable the formalization of narrative guideline content into care processes.

The development of new tools for intelligent information technology-based text analysis make it possible to access knowledge out of unstructured text and use it, for example, for decision support or research purposes. Most of the available tools as well as terminologies are available in English and cannot be used directly for German text or guidelines. Nonetheless, there are various international approaches. One promising approach is machine learning natural language processing with information extraction techniques [8]. Extracting information from clinical texts has been the focus of a growing body of research in the past few years. One of the fundamental tasks in clinical natural language processing research is to extract clinically relevant entities (e.g. diseases and drugs) by using semantic standards such as Concept Unique Identifier (CUI) defined in the UMLS. One of the most proven natural language processing tools is the open-source natural language processing system for the extraction of information from electronic medical record clinical free-text Apache cTAKES [9]. Apache cTAKES already offers a variety of algorithms for text analysis and information extraction. The system was deployed at the Mayo Clinic and is currently an integral part of their clinical data management infrastructure, and has processed over 80 million clinical notes. It can be normalized to domain ontologies by using UMLS concepts.

Methods

Related Research and Dataset

Dr Katja Heiden developed a software tool called PathGuide for model-based generation of guideline-compliant pathways for use in different Hospital Information Systems [10]. The

classification is done on the basis of the elements of an HL7 meta-model and indicates the content of a narrative recommendation, e.g. medication, procedure, observation, and additional information. This step is supported by an integrated mark-up tool. For example, the guideline recommendation that ‘patients with early invasive breast cancer should have a baseline dual energy X-ray absorptiometry (DEXA) scan to assess bone mineral density’ is classified as an examination (observation). Figure 1 shows the integrated mark-up tool by PathGuide.

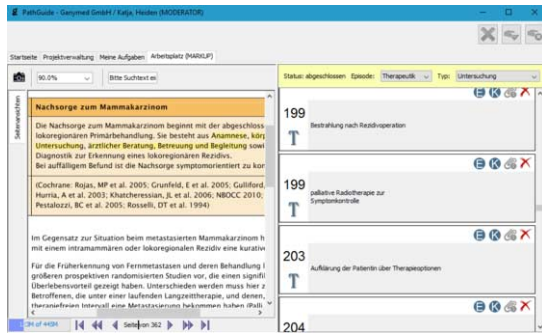


Figure 1-Mark-up Annotation by PathGuide [9]

The guideline and manually marked path-relevant passages are displayed on the left-hand side. On the right-hand side, a classification can be carried out manually. For this approach, the data model of PathGuide was extended by UMLS and the mark-up was supported by a natural language processing pipeline. For evaluation of the pipeline, an unstructured German breast cancer guideline was used.

Natural Language Processing

Natural language processing is an area of research and application that discovers how computers can be used to understand and manipulate natural language text or speech. The most common approaches to natural language processing are based on machine learning, a type of artificial intelligence that examines and uses patterns in data to improve a program’s own understanding. This research develops a clinical disorder recognition and encoding system by combining a machine learning-based approach for entity recognition with UMLS concept-mapping for German guidelines. One of the most proven natural language-processing tools is the open-source natural language processing system for extraction of information from electronic medical record clinical free-text Apache cTAKES [11]. It currently offers a variety of algorithms for text analysis and information extraction. With major modifications, such as the implementation of the German UMLS database and German OpenNLP models, it can be normalized to domain ontologies (such as SNOMED-CT or ICD 10) by using German UMLS concepts [12]. For pre-processing of German notes, three tiger data trained OpenNLP models were integrated into the natural language processing pipeline. These included a German Maxent Part-of-Speech tagger, Tokenizer and Sentence Detector. For the extraction of UMLS concepts from German clinical guidelines, a natural language processing pipeline with a mapping to the UMLS database is necessary [13]. Once the concepts are identified, it is possible to map the concepts to domain ontologies and international terminologies. In order to process the clinical guideline, a pre-processing step has to convert the guideline from PDF to plaintext.

Terminologies

The UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote the creation of more effective and interoperable biomedical information systems and services, including electronic health records. It is a comprehensive list of biomedical terms for developing computer systems capable of understanding the specialized vocabulary used in biomedicine and healthcare. For the natural language processing pipeline, the following German concepts from the UMLS database were integrated into the cTAKES database:

- International Statistical Classification of Diseases and Related Health Problems (ICD 10)
- Universal Medical Device Nomenclature System (UMDNS)
- The International Classification of Primary Care (ICPC)
- Logical Observation Identifiers Names and Codes (LOINC)
- Medical Dictionary for Regulatory Activities (MDR)
- Medical Subject Headings (MeSH)
- World Health Organization—Adverse Drug Reaction Terminology (WHO-ART)

The UMLS was developed by the National Library of Medicine (NLM). NLM produces and distributes UMLS Knowledge Source databases and associated software tools for use by system developers to build or enhance electronic information systems that create, process, retrieve, integrate, and/or aggregate data and informatics research. All UMLS Knowledge Sources and associated software tools are free of charge to U.S. and international users. The German database contains over 200,000 concepts.

Evaluation

Precision and recall are the basic measures used in evaluating search strategies. In pattern recognition and information retrieval with binary classification, precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. The evaluation of the adapted pipeline follows the standard metrics of evaluation using F1 (3), i.e. the harmonic mean of recall (2) and precision (1).

$$P = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (1)$$

$$R = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (2)$$

$$F1 = \frac{2PR}{(P+R)} \quad (3)$$

Precision (P) is the number of correct positive results divided by the number of all positive results, and recall (R) is the number of correct positive results divided by the number of positive results that should have been returned. The F1 score can be interpreted as a weighted average of precision and recall, where an F1 score reaches its best value at 1 and worst value at 0.

Results

For semantic data visualisation of the semi-automatic results, the Unstructured Information Management Architecture (UIMA) Annotation Viewer adapted to the cTAKES type-system was used [14]. Figure 2 shows the mark-up annotation results:

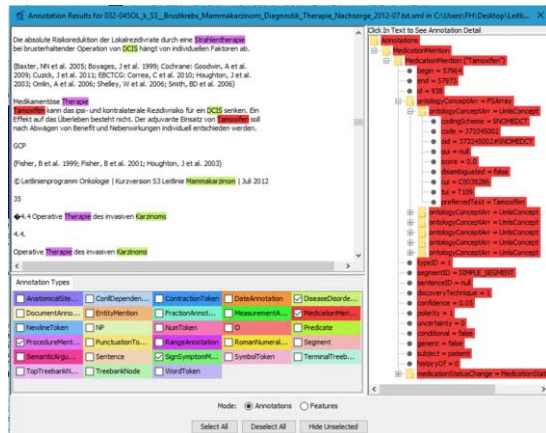


Figure 2-Mark-up Annotation Results

The Annotation Viewer on the left-hand side shows the semi-automatic mark-up by the natural language processing pipeline. Unlike PathGuide, the Annotation Viewer uses different colours to represent different Annotation Types (e.g. Diseases, Medication, Symptoms, and Anatomical Mention). On the right-hand side, the Annotation Viewer represents the identified and marked UMLS Concepts and the derived ontologies such as SNOMED-CT, ICD or RxNorm. The natural language pipeline only identifies the concepts that are also present in the UMLS database, without checking whether these are necessary for the creation of the clinical pathway. Table 1 shows the total amount of the manually identified and the semi-automatic identified items.

Table 1: Total Amount of Identified Items

Guideline Section	Manual Mark-up	Semi-automatic Mark-up (UMLS)
Diagnosis	17	84
Therapy	55	165
Rehab	8	29
Aftercare	19	45
Total	99	323

As noted, the total amount of semi-automatic mark-up and annotated items are significantly more. As discussed above, this is due to the missing check to determine if this item is relevant to the pathway or not. In manual mark-up, the areas that are not used for the creation of the clinical pathway are not marked. This ability is missing from the natural language pipeline.

To find out whether the manually marked areas were identified by the pipeline and were additionally annotated with UMLS, the respective items were analysed. Table 2 shows the results of the semi-automatic mark-up, with outgoing of the manually marked result as the gold standard.

Table 2: Evaluation of semi-automatic mark-up

Guideline Section	Precision	Recall	F1
Diagnosis	0.16	0.77	0.28
Therapy	0.27	0.81	0.41
Rehab	0.17	0.63	0.27
Aftercare	0.22	0.53	0.31
Total	0.21	0.68	0.31

Results showed that 68% of the manually marked items could be annotated by the pipeline through UMLS. The pipeline also obtained a good recall value, especially in diagnosis and

therapy. Due to the lack of knowledge about the relevance of items for the creation of a clinical pathway, significantly more items were marked and annotated by UMLS concepts. Consequently, the results for the precision are significantly lower since more false-positive entries are identified by the pipeline. However, the annotation by UMLS is automatically possible and the semi-automatic mark-up can be used as support for creating a clinical pathway.

Discussion

It is important to identify a few study limitations. First, the clinical guideline was created from manually identified items and considered as the gold standard, but this might not be the best process for this guideline. Formalizing a guideline and identifying areas of importance for the clinical pathway are not always clear and can depend on the person doing the formalization. Some of the mark-ups considered as erroneous as compared to the reference could, in fact, come from clinical interpretation. Second, lack of a knowledge database, which is needed for the creation of the clinical pathway, greatly limited the level of precision. In particular, only the words for which there is an entry in the German UMLS database can be annotated. An example of this is the phrase, *patient information*. This is an important step in the clinical pathway and has to be marked in the guideline, but this word exists only in the English version of the UMLS and not as the translated German word *Patientenaufklärung*. This can be attributed to the fact that the English UMLS database (5,571,374 entries) is much more extensive than the German database (217,672 entries). As a result, a learning program must be used to train the currently referenced database with guideline-specific abilities. The challenge is to mark the trained extensions in such a way that an update of the UMLS database is still possible without overwriting new trained entries. This marker is not provided by the cTAKES database. In order to avoid overwriting the new trained entries with an update, it may be possible to provide them with a postfix or prefix. These entries would then have to be excluded when updating the UMLS database. Additionally, there are areas in the guideline that do not fit in to any UMLS concept, regardless of language. For this reason, the next step is to implement a self-learning UMLS database for the German language to improve the results of the pipeline.

An alternative approach would be to formalize the guideline manually without using natural language processing support as mentioned earlier and then annotating the resulting files (e.g. XML files) using natural language processing. In this case, the pre-mark-up would be omitted, but false-positive markings would not arise. It would retrofit an automatic UMLS annotation without the advantage of a semi-automatic mark-up.

In the next development step, the UIMA Annotation Viewer is expanded with a manual mark-up function including a connection to the cTAKES database. This allows the user to insert unmarked words directly into the cTAKES database so that they are identified at the next mark-up. Lastly, there will be Dummy CUI codes for words that are important for the creation of the clinical pathway. This does not affect the annotation of UMLS codes, but these words will be marked as clinical pathway-relevant in the future.

Conclusion

Natural language processing can be used as a pre-processing step that can simplify the mark-up process and UMLS annotation in clinical guidelines to create clinical pathways.

However, a manual mark-up and a manual post-processing cannot be replaced by that pipeline.

References

- [1] K. Vanhaecht, The impact of clinical pathways on the organisation of care processes, *PhD thesis. University Leuven, Faculty of Medicine*, 2007.
- [2] R. Lenz, M. Reichert, IT Support for Healthcare Processes - Premises, Challenges, Perspectives, *Data Knowl Eng* **61** (2007), 39–58.
- [3] K. Heiden, B. Böckmann, Structured knowledge acquisition for defining guideline-compliant pathways, *Stud Health Technol Inform* **186** (2013), 73–77.
- [4] K. Bernstein, and U. Andersen. Managing care pathways combining SNOMED CT, archetypes and an electronic guideline system, *Studies in Health Technology and Informatics* **136** (2008), 353–358.
- [5] B.T. McInnes, T. Pedersen, J. Carlis, Using UMLS Concept Unique Identifiers (CUIs) for word sense disambiguation in the biomedical domain. *AMIA Annual Symposium Proceedings*, Volume 2007, 533.
- [6] S. Bloehdorn, P. Haase, Z. Huang, Y. Sure, J. Völker, F. van Harmelen, R. Studer, Ontology Management. In *Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies*, 1st edition, *Berlin Heidelberg: Springer*; 2009.
- [7] R. Lenz, M. Reichert, IT Support for Healthcare Processes - Premises, Challenges, Perspectives. *Data Knowl Eng*, **61** (2007), 39–58.
- [8] J. Franke, G. Nakhaeizadeh, I. Renz, XML Retrieval and Information Extraction, in *Text Mining – Theoretical Aspects and Applications*, *Physica-Verlag, Germany*, 2003, 29–32.
- [9] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association* **17** (2010), 507–13.
- [10] B. Böckmann, K. Heiden, PathGuide - model-based generation of guideline-compliant pathways for the use in different hospital information systems, *Stud Health Technol Inform* **192** (2013), 1089.
- [11] G. Divita, Q. Zeng., A. Gundlapalli, S. Duvall., J. Nebeker, M. Samore., Sophia, A expedient UMLS concept extraction annotator, *Journal of the American Medical Informatics Association* (2014), 467–476.
- [12] T. Benson, Clinical Terminology, Principles of Health Interoperability HL7 and SNOMED (Health Information Technology Standards), *Springer*, 2012.
- [13] M. Becker, B. Böckmann, Extraction of UMLS® Concepts Using Apache cTAKES™ for German Language, *Stud Health Technol Inform*, **223** (2016), 71–76.
- [14] UIMA, Apache UIMA, <https://uima.apache.org/downloads/releaseDocs/2.2.2-incubating/docs/html/tools/tools.html>, last access: 28.11.2016.

Address for correspondence

Matthias Becker, University of Applied Sciences and Arts, Department of Medical Informatics, Emil-Figge-Str. 42, 44227 Dortmund, Germany, Matthias.Becker@FH-Dortmund.de

Inter-Annotator Agreement and the Upper Limit on Machine Performance: Evidence from Biomedical Natural Language Processing

Mayla Boguslav and Kevin Bretonnel Cohen

Computational Bioscience Program, University Colorado School of Medicine, Aurora, CO, USA

Abstract

Human-annotated data is a fundamental part of natural language processing system development and evaluation. The quality of that data is typically assessed by calculating the agreement between the annotators. It is widely assumed that this agreement between annotators is the upper limit on system performance in natural language processing: if humans can't agree with each other about the classification more than some percentage of the time, we don't expect a computer to do any better. We trace the logical positivist roots of the motivation for measuring inter-annotator agreement, demonstrate the prevalence of the widely-held assumption about the relationship between inter-annotator agreement and system performance, and present data that suggest that inter-annotator agreement is not, in fact, an upper bound on language processing system performance.

Keywords:

Natural Language Processing; Supervised Machine Learning; Evaluation Studies

Introduction

The Code of Ethics and Professional Conduct of the Association for Computing Machinery includes the imperative to share knowledge of the *limitations* of computer systems (ACM Code of Ethics and Professional Conduct 2.7) [1]. In natural language processing, human-annotated data is the gold standard for most evaluation studies [2], and therefore it is crucial for understanding the limits of our work. It is standard practice to measure the quality of that data by assessing the extent to which humans agree with each other in the task of producing it [3]. This is called inter-annotator agreement (IAA) [4]. A standard assumption in the field is that the inter-annotator agreement establishes an upper bound on system performance [5-10]. In fact, the assumption that it is an upper bound on system performance turns out to be just that—a heretofore-untested assumption. The goal of the work reported here is to test that assumption. We do so by searching for the basis of that assumption; demonstrating that it is, in fact, a widely held assumption; and then collecting published findings in which system performance has exceeded the inter-annotator agreement and building simple statistical models of their relationships. This is important because if the assumption turns out not to be supported, as a community, we may be mis-estimating the actual performance of our systems. In particular, we may be over-estimating the quality of their performance by under-estimating how good it could potentially be.

Background

The calculation of inter-annotator agreement (often known outside of corpus linguistics as *inter-rater agreement*) is motivated by the need to deal with the problem of subjectivity in judgments about things that are not observable with the senses, a classic case of this being semantics. Lenaars [11] traces its roots back to logical positivism, and Krippendorff brought it to linguistic data in particular in the context of content analysis [12]. Exactly how inter-annotator agreement should be calculated remains an open topic of discussion.

Cohen [13], focusing on research in psychology, proposed quantifying the reproducibility and reliability of categorization by calculating the agreement between two annotators and correcting it for the probability of agreement by chance. This measure is known as Cohen's Kappa:

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

...where $Pr(a)$ is the observed agreement between two annotators and $Pr(e)$ as the expected agreement between the annotators if each annotator randomly picked a category for each annotation. Thus, Cohen's Kappa adjusts for chance to determine how much better the annotators did than chance [13]. Typically, language processing researchers compare the IAA score to the F_1 measure obtained by the system, including all of the papers discussed here except for [14], which uses precision (positive predictive value).

The F_1 is the harmonic mean of precision (P) and recall (R) (sensitivity). It is calculated on the basis of the numbers of true positives, false positives, and false negatives in a system's output:

$$F_1 = \frac{2PR}{P + R}, F_\alpha = \frac{(\alpha + 1)PR}{\alpha P + R}$$

In the case of annotating linguistic data, it is often the case that the expected chance agreement ($Pr(e)$ in the formula for kappa) is effectively zero, since there is no clear definition of what would count as a false positive, e.g. in the case of any task that requires the labelling of boundaries, such as in named entity recognition or any task involving scope (e.g. syntactic analysis). When this is the case, kappa is equivalent to F-measure, and this observation is the justification for their comparison here [15].

Methods

This paper approaches the assumption of inter-annotator agreement as the upper limit on system performance in three steps. First, we seek to answer the question of whether it is, indeed, a widely held assumption in the natural language processing community. Then, we try to find the source for this

assumption—the definitive citation. Finally, we look for counter-examples to the assumption; having found some, we do a statistical analysis of the papers that report results that contradict the assumption, on the rationale that if those results are just “noise,” it should show up in the descriptive statistics.

Since we are trying to characterize the community’s shared assumptions and to find the source of those assumptions, we took the literature as a proxy for those assumptions. We did a search for papers that explicitly assert that inter-annotator agreement is an upper bound on machine performance, and we looked for the sources that are cited in support of that claim. We carried out two separate searches. One was done by a natural language processing researcher. The other was done by a literature search service. We had worked with them in the past, and knew them to be quite competent in researching questions related to natural language processing. The full instructions given to the literature search service are available on this project’s GitHub site [16]. Briefly, they were as follows:

1. Find papers that assert explicitly that inter-annotator agreement is an upper bound on system performance.
2. Identify the source citation for that assertion.

We then asked the literature search service to find examples of papers that reported inter-annotator agreement *and* results from a natural language processing system, such that the system performed higher than the inter-annotator agreement.

To search the full text of publications, the service used Google Scholar. Phrasal search for *inter-annotator agreement* and *F-measure* and proximity operators to find cases where they occur near each other were used to retrieve an initial set of around 100 papers. Those papers were then examined manually, and any papers in which the inter-annotator agreement was higher than system performance *or* there was no explicit discussion of the relationship between them were excluded. This resulted in a set of 6 papers that included data on 20 systems that outperformed the inter-annotator agreement.

We next extracted the IAA and system performance measure for all 20 systems described within those articles. To evaluate the possibility that these values were noise, rather than an actual finding, we used simple statistical models to test for structure in the relation between IAA and system performance in three data sets: the systems that outperformed the IAA, other systems that did not, and both combined. The reasoning here is that if the findings are noise, that should be reflected as random variation in the F-measure, the IAA, or both; on the other hand, if it is not just noise, that would be reflected by structured relation.

For all three datasets, we used the Shapiro-Wilk test [17] to determine if they were normally distributed. We calculated the correlation between IAA and F-measure, reasoning that if the

papers that report outperforming IAA are just observing noise, there should be no relationship between them. Because most of the distributions were not normal, Spearman’s correlation, a non-parametric test [18], was used to calculate the correlations. The details are available on the GitHub site [16].

Results

There is no citation that establishes inter-annotator agreement as an upper bound on system performance. Neither we nor a professional literature search service found an authoritative citation for the idea that inter-annotator agreement is the upper bound on language processing system performance. It is often asserted, but we have not found a cited source that establishes it to be the case. None of the papers that explicitly asserted the assumption cited a source for the assertion.

Nonetheless, explicit statements of the assumption in multiple papers demonstrate that this assumption is widespread. We give six explicit statements of the assumption, including in papers by some of the most prominent researchers in the field—see the quotes in Table 1.

Table 1 – Explicit statements of the assumption of IAA as an upper bound in the natural language processing literature

Paper	Quote
Resnik and Lin [8]	“It is generally agreed that human inter-annotator agreement defines the upper limit on our ability to measure automated performance”
Gale, Church, and Yarowsky [10]	“An estimate of the upper bound is obtained by assuming that our ability to measure performance is largely limited by our ability to obtain reliable judgements from human informants”
Ormandjieva, Hussain, and Kosseim [7]	“...the average inter-annotator agreement...should be seen as upper bounds on the accuracy of any classifier”
Navigli [6]	“[Inter-annotator agreement] numbers lead us to believe that a credible upper bound for unrestricted find-grained word sense disambiguation is...”
Meyer and Gurevych [5]	“Besides the inter-annotator agreement A–B, which serves as an upper bound...”
Padó and Lapata [9]	“...the upper bound given by the inter-annotator agreement on the calibration data set”

Six papers (four from the biomedical domain and two from the general domain) reported at least one system that outperformed the IAA (see Table 2), for a total of 20 systems. Note that generally system performance was measured using F_1 measure, so we will use those terms interchangeably. The small number of papers reflects the fact that this is not a commonly reported phenomenon. However, neither is it unattested—this was not just a single counter-example, and those six papers reported on 20 systems that outperformed the inter-annotator agreement.

*Note that all articles use F_1 for system performance except for [14], which uses precision.

Table 2 – Systems that outperform the IAA

Paper	Systems that outperformed the IAA
<i>Combining Terminology Resources and Statistical Methods for Entity Recognition: an Evaluation</i> [19]	<ul style="list-style-type: none"> Machine learning to recognize specific entities within clinical notes Classifying intervention had lowest IAA and $F_1 \geq$ IAA with multiple methods
<i>Disambiguation of Occurrences of Reformulation Markers</i> [14]	<ul style="list-style-type: none"> Reformulation vs. non-reformulation in French with specific markers ESLO1/2 (spoken scenarios): Precision* > IAA
<i>SemEval-2015 Task 6: Clinical TempEval</i> [20]	<ul style="list-style-type: none"> Multiple Systems compete to identify critical timeline components of clinical notes and pathology reports from the Mayo Clinic Adj-Ann: IAA between adjudicator (final judge of the data to generally be used to train the system) and 2 annotators Many systems $F_1 \geq$ IAA and a few better than Adj-Ann (stronger)

Paper	Systems that outperformed the IAA
<i>Automatically Detecting Acute Myocardial Infarction (AMI) Events from EHR Text: a Preliminary Study</i> [21]	<ul style="list-style-type: none"> Automate the annotation of Worcester Heart Attack Study for AMI F₁ of system for ICD Diagnosis outperformed the IAA
<i>Deception Detection using Real-Life Trial Data</i> [22]	<ul style="list-style-type: none"> Deception detection System performance using decision trees significantly higher than annotator agreement and kappa statistic (0.01-0.20) Humans detect deception only slightly above chance
<i>Automatic Classification of Lexical Stress Errors for German CAPT (Computer-Assisted Pronunciation Trainings)</i> [23]	<ul style="list-style-type: none"> Classify non-native German lexical stress errors from manually annotated corpus of German word utterances by native French speakers IAA only fair

Figure 1 shows the F1 and inter-annotator agreement for the 20 systems. The Shapiro-Wilk normality test [17] showed that only the system performance measure is normally distributed. IAA is not, and skewed left. Therefore, we calculated the Spearman correlation, which is non-parametric. This showed that IAA and F₁ measure are significantly positively correlated ($\rho = 0.807$, $p\text{-value} = 8.56 \times 10^{-6}$) (see Figure 2).

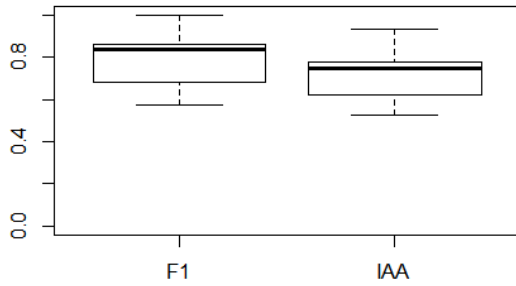


Figure 1 – Boxplot of F₁ measure (system performance) and IAA (annotator agreement) for systems that outperform the IAA

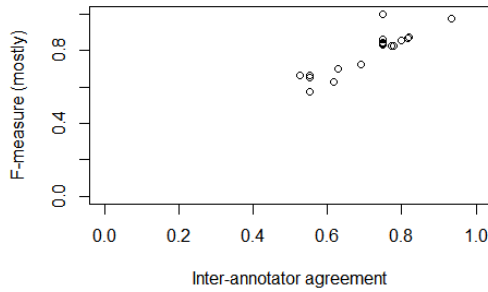


Figure 2 – Positive correlation between system performance and inter-annotator agreement for systems that outperform the IAA.

We then did the same analysis both for systems that did not beat the IAA, and for all systems together. In systems that did not outperform the IAA, neither IAA nor F-measure were normally distributed. There was a significant positive correlation between IAA and system performance ($\rho = 0.653$, $p\text{-value} = 1.449 \times 10^{-11}$) (see Figure 3).

For the combined data combined for systems that did and did not outperform the IAA, the IAA and system performance were significantly positively correlated, but less so compared to only the systems that outperformed the IAA ($\rho = 0.513$, $p\text{-value} = 1.81 \times 10^{-8}$) (see Figure 4).

We can also see how the medians are affected depending on which systems are included: those that outperform the IAA, those that do not, or both (see Table 3).

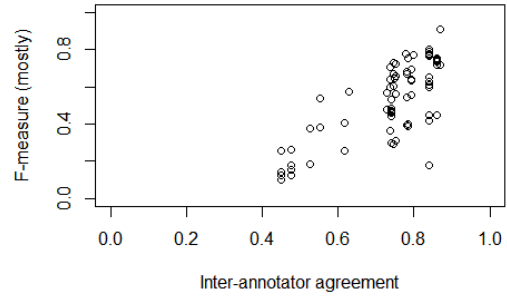


Figure 3 – Positive correlation between F₁ measure (system performance) and IAA (annotator agreement) for systems that do not outperform the IAA.

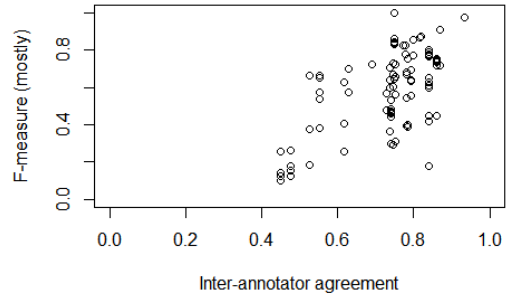


Figure 4 – Positive correlation between the IAA and F₁ measure for all data combined (both systems that do and do not outperform IAA).

Comparing the relationships between inter-annotator agreement and F-measure in the three sets of systems—ones which did outperform the IAA, ones which did not outperform the IAA, and the combination of those two, the relationships were the same—significantly positively correlated. This similarity across the three groups suggests that the cases of outperforming the inter-annotator agreement are not just noise.

Table 3 – Medians of IAA, system performance, and their difference across all data sets

Data	Median IAA	Median performance	Median difference
System > IAA	0.75	0.836	0.0785
System < IAA	0.7647	0.5865	-0.1655
All systems	0.7504	0.6380	-0.1383

Discussion

Providing accurate information to the public about technology and research results—as well as making funding decisions—requires the ability to accurately interpret measures of performance. The data presented here shows that a common standard for assessing natural language processing tools may overestimate their performance: contrary to a widely-shared and hitherto unexamined assumption in the field, the inter-annotator agreement is not necessarily the upper bound on performance in natural language processing. A lack of awareness of this can lead to the belief that systems are performing as well as they can, when in fact they are not.

Based on two literature searches and on the fact that no one ever cites one, there is no authoritative citation for the idea that IAA is the upper bound on system performance – in fact, it has been only an untested assumption. If there is some authoritative source that establishes this, not only have we not been able to find it, but also apparently no one else has, either, since no one cites one.

Despite being an untested assumption, it is nonetheless a widely-held assumption, as shown by the fact that we did not have any problem finding multiple explicit statements of it, some of them by the top people in our field.

Finally, the studies whose results are analyzed in this paper demonstrate that this widely-held assumption is not true. The distributional characteristics of the results—that is, their correlated, rather than unstructured nature—suggest that this is a real phenomenon, and not just noise. We cannot assume that the inter-annotator agreement is an upper bound on system performance, and in doing so, we may be over-stating how good natural language processing systems are.

In some of the papers that reported performance better than the inter-annotator agreement, the authors pointed out explicitly the unusualness of that finding. Some saw this as needing explanation, and they suggested explanations that were consistent with typical assumptions about inter-annotator agreement, such as that low inter-annotator agreement reflects a poor problem definition, an inherently difficult problem, poor guidelines, or—commonly—poor annotators [10; 20; 23-25].

If inter-annotator agreement does not establish the upper bound for system performance, what *should* we use in estimating the upper bound on system performance? Although a definitive answer to that question is outside of the scope of this paper, we discuss three possible solutions. They are based on changing what metric we use to quantify agreement; on changing who we define as the raters between whom the agreement is being calculated; and on replacing the agreement altogether with probabilistic estimates of a label quality.

One possibility is that we can safely use an inter-annotator agreement if we calculate it as something other than kappa. Although kappa is the most commonly reported measure of inter-annotator agreement, it has a number of problems. Some of these are essentially cultural—although there are a number of ways to calculate the expected chance agreement that is at the core of its claimed advantages, authors rarely report how they calculated the expected chance agreement. Consequently, it is often unclear what the kappa number actually reflects. When combined with the fact that the sensitivity of kappa to the probability of an estimated chance agreement is unstable—above an estimated chance agreement of about 0.5, kappa is extremely sensitive to small changes in the probability of chance agreement, while being relatively insensitive to small changes in the probability of a chance agreement below that

value—it is clear that there are many reasons to be suspicious of reliance on this number.

While in the previous paragraph we have discussed calculating something other than kappa to characterize the inter-annotator agreement, Bethard et al. [20] suggest changing the definition of the raters, such that rather than calculating agreement between two annotators, we calculate agreement between an annotator and an adjudicator. This may provide an agreement value that is more reflective of the data on which the system will be trained and evaluated, since if adjudicated data is available, that is typically what is used for training and testing. However, a number of conditions must be met for this to be possible—at minimum, there has to be an adjudication step, which is not always the case. Furthermore, changing the definition of the raters between whom agreement is calculated does not answer the question of how to calculate the agreement between them.

Finally, Passoneau and Carpenter suggest abandoning an agreement entirely and building a probabilistic annotation model of label quality [26].

In the larger context of responsible conduct of science, the findings reported here are relevant to the small but growing body of work on the ethics of NLP [27-29]. As noted above, the ethical standards of the Association for Computing Machinery include the responsibility to communicate the *limitations* of computer systems [1]. In reporting performance, there is a common assumption that metrics that approach an inter-annotator agreement reflect *high* performance [5-10]. The data reported here suggest that such performance may not be as high as we think it is, relative to the best possible performance, suggesting that this assumption can lead—certainly inadvertently—to conduct that does not meet the Association for Computing Machinery standards.

Conclusion

This paper examines a common assumption in natural language processing. It is shown that the assumption is, indeed, widespread; that there is no established justification for that assumption; and that the assumption is not true. This last point is demonstrated both by multiple counterexamples, and by descriptive statistics that suggest that the counterexamples are not random noise in the larger population of published papers on language processing, but rather reflect a real phenomenon. Responsible conduct of science will be enhanced by being aware of this.

Acknowledgements

Boguslav is supported by the Dean's Fund at University of Colorado Anschutz Medical. Cohen is supported by NIH grants LM008111, LM009254, and NSF IIS-1207592 to Lawrence E. Hunter, and by generous funding from Labex DigiCosme (project ANR11LABEX0045 DIGICOSME) operated by ANR as part of the program «Investissement d'Avenir» Idex ParisSaclay (ANR11IDEX000302), as well as by a Jean d'Alembert fellowship. The work was aided by discussions with Patrick Paroubek, Bob Carpenter, and Tiffany Callahan; all remaining faults are the authors'.

References

- [1] R.E. Anderson, G. Engel, D. Gotterbarn, G.C. Hertlein, A. Hoffman, B. Jawer, D.G. Johnson, D.K. Lidtke, J.C. Little, D. Martin, D.B. Parker, J.A. Perrolle, and R.S. Rosenberg, ACM Code of Ethics and Professional Conduct, in, ACM, 1992.
- [2] P. Jackson and I. Moulinier, *Natural language processing for online applications: text retrieval, extraction, and categorization*, John Benjamins Publishing Company, Amsterdam, 2007.

- [3] J. Pustejovsky and A. Stubbs, *Natural language annotation for machine learning*, O'Reilly Media, Sebastopol, CA, 2013.
- [4] R. Artstein and M. Poesio, Inter-Coder Agreement for Computational Linguistics, *Association for Computational Linguistics* 34 (2008), 555-596.
- [5] C.M. Meyer and I. Gurevych, Worth its weight in gold or yet another resource—A comparative study of Wiktionary, OpenThesaurus and GermaNet, in: *Gelbukh A. (eds) Computational Linguistics and Intelligent Text Processing*, Springer, Berlin, Heidelberg, 2010.
- [6] R. Navigli, Meaningful clustering of senses helps boost word sense disambiguation performance, in: *International Conference on Computational Linguistics and the annual meeting of the Association for Computational Linguistics*, Sydney, Australia, 2006.
- [7] O. Ormandjieva, I. Hussain, and L. Kosseim, Toward a text classification system for the quality assessment of software requirements written in natural language, in: *Fourth International Workshop on Software Quality Assurance, SOQUA 2007, in conjunction with the 6th ESEC/FSE joint meeting*, Dubrovnik, Croatia, 2007.
- [8] P. Resnik and J. Lin, Evaluation of NLP Systems, in: *The handbook of computational linguistics and natural language processing*, C.F. Alexander Clark, Shalom Lappin, ed., Wiley-Blackwell, 2010, pp. 271-295.
- [9] S. Padó and M. Lapata, Cross-linguistic projection of role-semantic information, in: *HLT '05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Vancouver, British Columbia, Canada, 2005, pp. 859-866.
- [10] W. Gale, K.W. Church, and D. Yarowsky, Estimating upper and lower bounds on the performance of word-sense disambiguation programs, in: *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1992, pp. 249-256.
- [11] A.A. Leenaars, *Suicide Notes: Predictive Clues and Patterns*, Human Sciences Press, Inc., New York, 1988.
- [12] K. Krippendorff, *Content analysis : an introduction to its methodology*, Sage Publications, Beverly Hills, 1980.
- [13] J. Cohen, A Coefficient of Agreement for Nominal Scales, *Educational and Psychological Measurement* 20 (1960), 37-46.
- [14] N. Grabar and I. Eshkol-Taravella, Disambiguation of occurrences of reformulation markers c'est-à-dire, disons, ça veut dire, *JADT 2016* 7 (2016).
- [15] G. Hripcsak and A.S. Rothschild, Agreement, the f-measure, and reliability in information retrieval, *J Am Med Inform Assoc* 12 (2005), 296-298.
- [16] K.B. Cohen and M. Boguslav, KevinBretonnelCohen/InterAnnotatorAgreement, in, GitHub, GitHub Inc., 2016.
- [17] S.S. Shapiro and M.B. Wilk, An Analysis of Variance Test for Normality (complete Samples), *Biometrika* 52 (1965), 591-611.
- [18] M. Mukaka, A guide to appropriate use of Correlation co-efficient in medical research, *Malawi Medical Journal: The Journal of Medical Association of Malawi* 24 (2012), 69-71.
- [19] A. Roberts, R. Gaizauskas, M. Hepple, and Y. Guo, Combining Terminology Resources and Statistical Methods for Entity Recognition: An Evaluation, *LREC: European Language Resources Association*. (2008), 2974-2980.
- [20] S. Bethard, L. Derczynski, G. Savova, J. Pustejovsky, and M. Verhagen, SemEval-2015 Task 6: Clinical TempEval, in: *SemEval 2015*, Association for Computational Linguistics, Denver, Colorado, 2015, pp. 806-814.
- [21] J. Zheng, J. Yarzebski, B.P. Ramesh, R.J. Goldberg, and H. Yu, Automatically Detecting Acute Myocardial Infarction Events from EHR Text: A Preliminary Study, *AMIA Annu Symp Proc* 2014 (2014), 1286-1293.
- [22] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, Deception Detection using Real-life Trial Data, in: *ICMI 2015*, New York, NY, USA., 2015, pp. 59-66.
- [23] J.T. Anjana Sofia Vakil, Automatic classification of lexical stress errors for German CAPT, in: *Workshop on Speech and Language Technology in Education (SLaTE 2015)*, Leipzig, 2015, pp. 47-52.
- [24] R.G. A. Roberts, M. Hepple, Y. Guo, Combining terminology resources and statistical methods for entity recognition: an evaluation, *LREC: European Language Resources Association* (2008), 2974-2980.
- [25] J. Zheng, J. Yarzebski, B.P. Ramesh, R.J. Goldberg, and H. Yu, Automatically Detecting Acute Myocardial Infarction Events from EHR Text: A Preliminary Study, in: *AMIA Annual Symposium*, AMIA, 2014, pp. 1286-1293.
- [26] R.J. Passonneau and B. Carpenter, The benefits of a model of annotation, *Transactions of the Association for Computational Linguistics* 2 (2014), 311-326.
- [27] K. Fort, G. Adda, and K.B. Cohen, Éthique et traitement automatique des langues et de la parole: truismes et tabous, *Traitement Automatique des Langues* 57 (2016), 7-19.
- [28] D. Hovy, S. Spruit, M. Mitchell, E.M. Bender, M. Strube, and H. Wallach, Ethics in natural language processing, *Association for Computational Linguistics* (2017).
- [29] D. Hovy and S.L. Spruit, The social impact of natural language processing, in: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2016.

Address for correspondence

kevin.cohen@gmail.com

Sharing Health Big Data for Research - A Design by Use Cases: The INSHARE Platform Approach

Guillaume Bouzille^{a,b,c,d}, Richard Westerlynck^{d,k}, Gautier Defossez^f, Dalel Bouslimi^{g,h}, Sahar Bayatⁱ,
Christine Riou^{a,b,d}, Yann Busnel^j, Clara Le Guillou^h, Jean-Michel Cauvin^h, Christian Jacquelin^t,
Patrick Pladys^c, Emmanuel Oger^e, Eric Stindel^h, Pierre Ingrand^f, Gouenou Coatrieux^{g,h},
Marc Cuggia^{a,b,c,d}

^a INSERM, U1099, - ^b Université de Rennes 1, LTSI, Rennes, 35000, France

^c CIC Inserm 1414, ^d Centre de Données Clinique, ^e Équipe CTAD-PEPI, CHU de Rennes, 35000, Rennes, France.

^f Registre général des cancers de la région Poitou-Charentes, Poitiers, France.

^g Institut Mines-Télécom; Télécom Bretagne, ^h INSERM 650 Latim, Brest, France

ⁱ EHESP Rennes, Sorbonne Paris Cité, France,

^j Agence de la biomédecine, Saint Denis, France, ^k CREST(ENSAI) / Inria Rennes, France

Abstract

Sharing and exploiting Health Big Data (HBD) allow tackling challenges: data protection/governance taking into account legal, ethical, and deontological aspects enables trust, transparent and win-win relationship between researchers, citizens, and data providers. Lack of interoperability: compartmentalized and syntactically/semantically heterogeneous data. INSHARE project using experimental proof of concept explores how recent technologies overcome such issues. Using 6 data providers, platform is designed via 3 steps to: (1) analyze use cases, needs, and requirements; (2) define data sharing governance, secure access to platform; and (3) define platform specifications. Three use cases – from 5 studies and 11 data sources – were analyzed for platform design. Governance derived from SCANNER model was adapted to data sharing. Platform architecture integrates: data repository and hosting, semantic integration services, data processing, aggregate computing, data quality and integrity monitoring, Id linking, multisource query builder, visualization and data export services, data governance, study management service and security including data watermarking.

Keywords:

Information Dissemination, Information Storage and Retrieval, Registries

Introduction

Health Big Data (HBD) is more than just a very large amount of data or a large number of data sources. It also refers to the complexity, challenges, and new opportunities presented by combined analysis of data. Health data collected or produced are now potentially sharable and reusable. They can be exploited at different levels and across different domains, especially concerning questions related to multidisciplinary research. This huge amount of data holds the promise of supporting a wide range of medical and health care functions, including among others clinical decision support, disease surveillance or population health management [1]. This explains the incentive policy of opening HBD around health data science being supported by different public authorities and

scientific communities such as OpenData, AVIESAN or Inserm initiatives, as well as European research programs like IMI or Horizon 2020. Recently, strong initiatives have been launched in U.S to enhance utility of health Big Data and finally to enter in the next level of knowledge discovery [2].

In this context, clinical data warehouse (CDW) technology comes forward as one of the solutions to address HBD exploitation. CDW, are becoming increasingly widespread in U.S, being put to use for different purposes including cohort discovery, biomarker detection, feasibility studies or enrolment of patients in clinical trials. Research communities are currently connecting CDW to one another with the aim of creating Clinical Data Research Networks (e.g. PCORNET [2]) or biomedical research network (e.g. Data to Knowledge).

In these networks, data providers such as researchers, health facilities, research agencies or institutions make part of their data available to research community while maintaining data sharing control at all time. Thus, these trusted third-party platforms integrate and open-up scientific or potentially scientific health data [3]. This makes use of these data at a large-scale possible. In France, such platform, that would be able to integrate and share multisource and multiscale big and small health data produced by health institutions for research purposes, does not exist.

This is the aim of the INSHARE French national project; in which different and actual key issues like governance, organizational, and technical factors to perform such data sharing will be explored and addressed. The absolute goal is to facilitate access to data and foster collaborative research and data sharing between researchers and data providers. In this paper, we present and discuss these key issues and approach we are following to design the platform. The approach is driven by real research use cases of high interest for individuals and states.

Background

Data to share: In its large acceptance HBD sources comprise various types of data from structured information such as OMICS data, administrative or billing data, drug prescription data consisting of dates and dosages captured through

standardized ePrescription system, to unstructured and textual data such as clinical narratives that describe medical reasoning behind prescriptions [4]. Beyond data generated by hospitals, several health data sources come from health registries or insurance databases, which are a valuable source of standardized, longitudinal, population-wide data. For instance, the French health reimbursement database (Système National d'Information Inter-Régimes de l'Assurance Maladie, SNIIR-AM) contains individualized, anonymous and comprehensive data for all health spending reimbursements received by affiliated subjects, including basic patient demographic data such as age, gender, medical drugs, and outpatient medical cares – prescribed or performed by health professionals from both public and private practices. SNIIR-AM is also linked via a unique personal health number to the French hospital discharge database (PMSI), which contains diagnostic codes, medical procedures, and admission dates for all hospitalizations. Data from SNIIR-AM is increasingly used for research projects, especially relating to detection of drugs' adverse effects in epidemiology or clinical research. The designed platform governance models were derived from the SCANNER model and adapted for data sharing.

Sharing barriers: The regulatory hurdles obstructing optimal use of data for research have been extensively discussed within specialised literature [5]. Identified factors are characterized by (i) over-cautious approach among data custodians, many of whom are unwilling to link or share data, (ii) legislators' failure to consider flexibility required to allow and support such linking and sharing and (iii) incorporation of 'good governance' models or intelligent design of working instances not contemplated within the regulatory framework, nor reflecting on the subject [5], [6]. Sethi proposes a model for data sharing governance including (i) guiding principles and best practices, (ii) safe, effective and proportionate governance, (iii) articulation of roles and responsibilities of data controllers and data processors and (iv) development of a training program for researchers that covers appropriate vetting procedures prior to sharing valuable data.

Cornerstone of data sharing and reuse is trust. Therefore, implementing a trustworthy process for handling citizens' and patients' health data is a pivotal goal. Based on the definition of a trusted relationship, one party (trustor) is willing to rely on the actions of another party. In addition, the trustor abandons (full) control over the actions performed by the trustee. As a consequence, a trustworthy system is that in which, the trustor can "place his/her trust and rest assured that the trust will not be betrayed". A system for data reuse should thus prove its trustworthiness by fulfilling the responsibility of dealing with data within the limits of a social contract regulated by policies between citizens and organizations handling the system.

The technological components behind a trustworthy system involve designing and implementing IT tools and services capable of guaranteeing data quality and security while providing interoperability, adaptability, and scalability. Specific projects funded by the EU and by the IMI initiative [7], such as EHR4CR, are dealing with such challenges, with the prospect of defining use cases, tools, technologies and a business model for data reuse. In particular, the EHR4CR business model includes accreditation and certification plans for EHR systems that can be integrated within a system for data reuse. The purpose of data reuse has implications that belong to the realm of policies and regulations, which are essential aspects for establishing trust. How to manage informed consent is one of the key aspects connected to this issue. In fact, current regulations in many European countries, which are similar to the US, with the HIPAA act, assume that consent (implied or

explicit) for use of data is strictly limited to the purpose for which data were collected. This may seriously limit the scope of data analysis.

This theme needs to be reconsidered in the light of the existence of a proper, trustworthy system based on an agreement between citizens and healthcare organizations. Specific practical examples of policies for handling data reuse are provided by regional initiatives in Europe, two such cases being the United Kingdom and Catalonia. ISO/TS 14265:2011 provides a classification of different purposes for processing personal health information that can help make policy formulation more granular.

Methods

To design organizational and technical dimension of the INSHARE platform, an iterative and 4 step bottom-to-top approach has been adopted, by analyzing on the ground, existing needs, use cases, and actual difficulties encountered by the project partners. Four partners are involved in the project as data providers: 2 academic hospitals (CHU Rennes and Brest) which provide datamarts from their Clinical Data Warehouse (eHOP-CDW), 3 epidemiologic registries at a regional or national scale.

This approach aims at defining technical and functional specifications, data protection policies and governance for an efficient and valuable data sharing. Furthermore, this approach takes into account the fact that some technological issues have to be addressed and especially the evolution needs for data analysis and security tools in the scaling-up to HBD.

Step 1 - To describe use cases and user needs:

The aim of this step is to define precisely scenarios from an operational perspective, the information workflow and system/actor interfaces that relate to exploitation of health and research data via the INSHARE platform. Relevant scenarios leverage the richness and variability of data sources hosted by the platform in terms of domain, quality, and origin. Herein, the objective is to identify the functional needs, which are expected by different users of the platform: researchers- users, data providers, and internal operators of the platform.

Step 2 - To define data sharing governance and secure access to the platform:

Regarding ethical, legal and deontological aspects, a focus group composed of domain experts and representatives of patient associations conducts this study. According to the state-of-the-art step and specified use cases defined at the first step, the objective is to establish governance guidelines guaranteeing data protection and individuals' privacy rights. This step includes submission of these guidelines for validation to institutional and regulatory authorities such as the Comité consultatif sur le traitement de l'information en matière de recherche (CCTIRS) and the Commission Nationale de l'Informatique et des Libertés (CNIL), two French authorities in charge of such regulation aspect.

Step 3 - To define INSHARE platform specifications:

The aim of this step is to define a comprehensive description of intended purpose and environment of the platform. These specifications describe what the software does and how it will be expected to perform, taking into accounts the operational scenarios, security aspects, and stakeholders (users, data providers, and data managers) inputs. It also addresses some key issues in relation with data analysis and security. In terms of data protection in the scaling up, special interest is given to data traceability and on how to give back some control to data providers on the data they make available to researchers. On one hand, users have to know of their action accountability and, in another hand, patient or data provider consent for data exploitation duration has to be guaranteed. Database watermarking, a very recent solution [8], is one of the technology actually explored for those purposes.

Each data provider is part of the INSHARE project to bring their knowledge and experience with their respective data. Data providers are thus responsible for supplying necessary data to the platform in order to answer to the use cases. They have to supply all necessary information about data to correctly perform their integration and to subsequently give capability to the platform to provide the best-suited data for each user request.

Results

Use Cases: Three main use cases corresponding to 5 studies and application domains have been identified and chosen to be performed on the platform. Being able to ensure one of them will be of great interest for cares of individuals and populations. Table 1 illustrates for each use case the sources of data, which will be shared and used in the different INSHARE platform studies.

Table 1 – Use cases and application domain

Use Case	Study and application domain	Data Sources
Health care Trajectory analysis	- Pre & post-dialysis care trajectory of end-stage renal disease patients starting dialysis in emergency - Characterizing the healthcare trajectories of children (and their mother) included in Birth Defect Registry	Kidney Failure registry (REIN) SNIIR-AM Birth defect registry eHOP-CDW
Registry enrichment	Assessment of association between cancer incidence and diabetes in end-stage renal disease patients	Kidney Failure registry (REIN) Cancer registry SNIIR-AM
Signal detection	- Influenza surveillance - Adverse drug effect surveillance	eHOP-CDW SNIIR-AM Sentinel Network Open Data

For instance, regarding study on health care trajectories of end-stage renal disease patients, comorbidities are currently

collected and registered in the REIN database at initiation of renal replacement therapy (RRT). But occurrence of comorbidities after RRT started is not a mandatory field of REIN [9]. Moreover, no information on prescribed treatments is available in the REIN database. Through the INSHARE platform, the hospital CDW (Rennes and Brest) will be used to collect comorbidities and expensive drug prescriptions while drug exposures will be extracted from SNIIR-AM in order to enrich the REIN registry with accurate comorbidities and medications including standards and dates of occurrence.

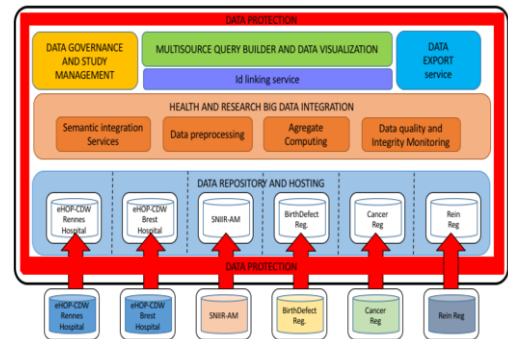


Figure 1 – Uses case under the scope of the Inshare Platform

Platform governance:

The governance determines how the range of controls and procedures with contractual obligations work together to ensure an end-to-end secure and trustfully platform, where the security and reliability of data is guaranteed. Indeed, several.

INSHARE use cases imply to perform Id linkage processing and require the access to identified data (e.g., Epidemiologic registry enrichment). Moreover, all use cases require aggregating data coming from multiscale institutions (local academic hospitals, regional to national registries, and for the SNIIR-AM, data issued from a nationwide database). Some of them imply intensive computation on big volume of data (e.g. signal detection). All these constraints have led to defining a model of governance for the platform adapted to big data sharing. The model we propose is derived and adapted from the Distributed Scalable National Network for Effectiveness Research (SCANNER). It consists of identifying 10 basic requirements: platform and data provider information, institution information, study information, ethical agreements (coming from an independent IRB-like comity), Data Sharing Agreement (from data providers who are involved in the study), approved users (external users and internal operators of the platform), authentication and access, data use, audit and accounting, patient rights, data segregation. In addition, according to the type of personal data, it includes de-identification, data watermarking, individual access, correction, openness/transparency, individual choice, use and disclosure limitation, integrity, accountability, and safeguards. Technology and some others can meet requirements by contracts, attestation of users, or management supervision.

Platform Design:

As a result, we designed the platform architecture shown in Figure 1. This architecture is oriented to meet the different use cases under the scope of the project, and to perform expected data processing while respecting governance framework mentioned above. The platform encompasses services of several weakly coupled components, the whole in a

cloud-oriented architecture. Hereby, we detail some of the key components and services:

Data repository and hosting component is a buffer zone where data providers make available required datasets or datamart to share. The core idea is to host in the platform data with the finest granularity and in their most original form, i.e., with the least transformation possible.

Health big data integration layer comprises components and services dedicated to data integration and processing. The semantic integration service (SIS) contains information models (i.e. database schemas of the data sources such as eHOP or SNIIR-AM) as well as semantic resources either used in the sources or required for semantic integration (reference or interface terminologies, ontologies and mappings) and ensures standards' interoperability (such as HL7, PN13, HPRIM). SIS provides tools and methods to the other standard components.

Data preprocessing service is devoted to data enrichment. It includes NLP tasks and data indexing to make easier extraction of useful information from large-scale data stored across the different INSHARE sources. A core functionality of these services is to execute data processing tasks and to query the data virtualization layer in order to access stored data. The developed engine is also responsible for planning, coordination, and execution of queries to the data virtualization layer in a distributed manner commercial data processing frameworks and parallel relational database management systems.

Aggregate computing service is designed for building online auxiliary indexing and summarization structures based on the incoming data processing tasks and their data requirements. Based on profiling and statistic information of the submitted processing tasks. For instance this service is used to compute from CPOE data, aggregates of drug dose per day, week, stay or globally for a patient or a population.

Data quality and integrity monitoring: The INSHARE platform deals with data sources having heterogeneous data quality, from EHRs to epidemiologic registries. Integration process has to manage such quality disparities. This component is dedicated to compute metrics to monitor data quality during the integration process. These metrics are useful to (i) alert data providers and take corrective actions at the data source, (ii) perform more accurate analysis taken into account possible bias due to data quality issues, (iii) improve data quality within the INSHARE platform, each source bringing complementary information. For instance, for the same patient, in and out hospital drug information come from different sources and is registered in different ways (structured and coded data from CPOE or SNIIR-AM, text for clinical charts, forms and notes). One source can provide more accurate or exhaustive information to others.

Id Linking service: For security reasons, in France, as in most countries, there are currently no patient identifiers that can be used to directly link data from different data sources. Nonetheless, several national programs or initiatives provide researchers either trusted third-party linkage services, or big, pre-linked datasets. This, for instance, is the case of the French hospital discharge database (used as part of the hospital billing system) that matches data coming from all hospitals in France. The SNIIR-AM is arguably one of the most noteworthy linked data sources recently opened to the research community. The Id Linking service reuses and provides methods to link data sources using deterministic and probabilistic approaches on common data elements. For instance DRG data coming from hospital are already linked with the other data of SNIIR-AM. EHOP CDW includes the DRG data. Even without specific

common Id, linking can be performed using dates, groups of diagnosis and procedure codes, and ADT mode.

Data Governance, study management service and security: These services encompass tools, procedures [10] and workflow to cover governance requirements and provide continuous data protection, from their acquisition to their outsourcing and mutualization within the INSHARE platform and beyond (e.g. when exported). The idea is to complement current data protection, which mainly relies on security of the information system and which do not make it possible to know if data are used for the purposes originally foreseen, especially when data are outsourced. The protection of digital content we deploy is based on watermarking [8] and crypto- watermarking solutions (i.e. mechanisms that combine encryption and watermarking [11] that fulfill different security objectives, in particular in terms of integrity and traceability (identification of information-leak sources or of end-user misbehavior). If data are protected as long as they are not decrypted, watermarking leaves free access to them and maintains them protected by means of security attributes (e.g. digital signatures, users' ID, access rights) invisibly inserted or embedded into the data themselves. Moreover, watermarking protection is independent of the data storage format. These data protection tools are designed to take into account strong interoperability constraints so as to: i) provide security resilient to information-processing; ii) make the protection on the data provider's side compliant with the one used by the INSHARE platform and beyond.

Multisource Query Builder, Visualization and data export services: These services are intended to: design and perform complex queries on multisource data; visualize results with different modalities; and, export processed dataset to the end users. From the end user point of view, interaction with the platform consists of submitting a request for a study to the platform. Only certified and authorized operators of the platform will have to access to the query workbench for data exploitation and eventually to export required datasets to the end user.

Figures 2 and 3 illustrate the workflow for two scenarios. In the first (Figure 2), a targeted research database is fed by data extracted both from two registries (REIN and Cancer playing the role of data provider) according a study protocol about the association between kidney failure and occurrence of cancer. This protocol, which defines criteria for data selection and variables to extract from the source, and user agreement are submitted to the platform. Figure 3 illustrates how the platform is able to enrich a data source (here the Rein registry) by collecting from a list of patient ID, missing or required data (e.g. comorbidities) from the different sources. In this scenario Rein registry is user of the platform and recipient of data. All along this process, health data is maintained, secured by means of digital content protection tools, with a special interest for data traceability and audit trails.

Discussion and Conclusion

The INSHARE project's consortium made the choice to focus on some of today's crucial challenges, which, in our opinion, are still not resolved: data quality assessment for research purposes, scalability issues when integrating heterogeneous health "big data" or patient data privacy and data protection. Moreover, adoption of electronic health data is still an active

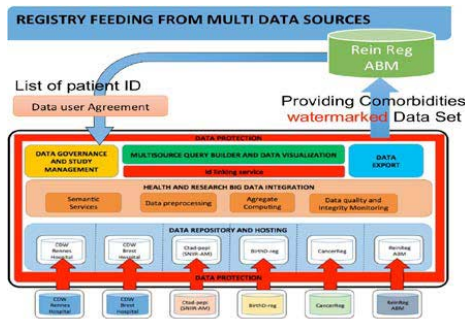


Figure 2 – Workflow for ad-hoc study and register enrichment

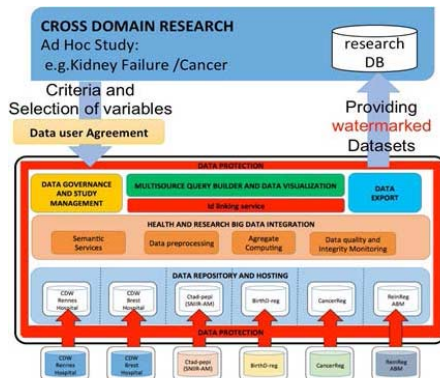


Figure 3 – Workflow for register enrichment

process and the way to exploit these data is rapidly evolving. In a second time, use of health data still requires developing to reach maturity: the aforementioned barriers have to be raised before we can obtain more significant knowledge and practical consequences. In any case, the project's use-cases were selected in view of their potentially important impact on clinical knowledge and health research, and we believe they could help demonstrate the benefits of secondary use of data. From a methodological point of view, our approach goes beyond today's most cutting-edge secondary-use technologies. We propose a combination of innovative algorithms for cryptography and watermarking supported by big data technologies, with the aim of enhancing content digital protection. Compared to other projects mainly oriented to distributed architecture, we also follow a different strategy based on a trusted third-party oriented on health-data sharing within a community of users.

Lead by the large amount of heterogeneous data available among the consortium, traditional approaches for data organization and analysis (using classical SQL server such as I2B2) are no longer efficient and quite overwhelmed. For instance, the Rennes CDW contains by itself raw data of 1.6 million patients. Recent Big Data technologies are filling the gap of this evolution that requires real time analysis, low latency, data mining, and heterogeneous unstructured data treatment. Born from needs of scalability, fault tolerance and interoperability, challenging frameworks come out as Hadoop or Cassandra. However, big data technologies are an evolving landscape. The INSHARE project is a great opportunity to test and evaluate combination of disruptive technologies and provide improved analysis performance in the perspective to

meet real-world use cases and users needs, on real and massive data. For instance, preliminary tests on adverse drug effect detection have been carried out. In this example, the combination of OrientDB (which is a graph oriented database) with SPARK for aggregate computing has shown promising performance for intensive computing. Nonetheless, applying existent solutions should not be sufficient in the background of the INSHARE project. Indeed, starting from the available massive datasets, a second objective aims to design innovative algorithms and techniques in a prospective way (using a data sciences approach, sharing the statistical and computer sciences skills

Acknowledgements

We would like to thank the French National Research Agency (ANR), for funding this work inside the INSHARE (INtegrating and Sharing Health dAta for Research) project (grant no. ANR-15-CE19-0024).

References

- [1] W. Raghupathi and V. Raghupathi, Big data analytics in healthcare: promise and potential, *Health Inf Sci Syst*, **2**(1) (2014), 3.
- [2] Patient-Centered Outcomes Research Institute (PCORI), PCORnet | National Patient-Centered Clinical Research Network. [Online]. Available: <http://www.pcori.net/>. [Accessed: 23-Apr-2015].
- [3] M. D. Natter, J. Quan, D.M. Ortiz et al., An i2b2-based, generalizable, open source, self-scaling chronic disease registry, *JAMIA*, **20**(1) (2013), 172-179.
- [4] P. B. Jensen, L. J. Jensen, and S. Brunak, Mining electronic health records: towards better research applications and clinical care, *Nat Rev Genet*, **13**(6) (2012), 395-405.
- [5] J. Peto, O. Fletcher, and C. Gilham, Data protection, informed consent, and research, *BMJ*, **328**(7447) (2004), 1029-1030.
- [6] N. Sethi and G. T. Laurie, Delivering proportionate governance in the era of eHealth, *Med Law Int*, **13**(2-3) (2013), 168-204.
- [7] M. Goldman, The Innovative Medicines Initiative: a European response to the innovation challenge, *Clin Pharmacol. Ther.*, **91**(3) (2012), 418-425.
- [8] J. Franco-Contreras and G. Coatrieux, Robust watermarking of relational databases with ontology-guided distortion control, *IEEE Trans. Inf. Forensics Secur.*, **10**(9) (2015), 1939-1952.
- [9] C. Couchoud, B. Stengel, P. Landais et al., The renal epidemiology and information network (REIN): a new registry for end-stage renal disease in France, *Nephrol Dial Transplant*, **21**(2) (2006), 411-418.
- [10] G. Bouzillé, E. Sylvestre, B. Campillo-Gimenez et al., An integrated workflow for secondary use of patient data for clinical research, *Stud Health Technol Inform*, **216**, (2014), 913.
- [11] D. Bouslimi and G. Coatrieux, A crypto-watermarking system for ensuring reliability control and traceability of medical images, *Signal Process Image Commun*, **47**, (2016), 160-169.

Address for correspondence

Pr Marc Cuggia,
UMR LTSI (Inserm), Faculté de médecine. Rue du Pr Léon
Bernard. 35043 Rennes Cedex. marc.cuggia@univ-rennes1.fr

Developing Visual Thinking in the Electronic Health Record

Andrew D. Boyd^a, Christine D. Young^a, Margret Amatayakul^b, Michael G Dieter^a, Lawrence M Pawola^a

^a Department of Biomedical and Health Information Sciences, University of Illinois at Chicago, Chicago, Illinois, U.S.A.

^b College of Applied Health Sciences Advisory Board, University of Illinois at Chicago, Chicago, Illinois, USA.

Abstract

The purpose of this vision paper is to identify how data visualization could transform healthcare. Electronic Health Records (EHRs) are maturing with new technology and tools being applied. Researchers are reaping the benefits of data visualization to better access compilations of EHR data for enhanced clinical research. Data visualization, while still primarily the domain of clinical researchers, is beginning to show promise for other stakeholders. A non-exhaustive review of the literature indicates that respective to the growth and development of the EHR, the maturity of data visualization in healthcare is in its infancy. Visual analytics has been only cursorily applied to healthcare. A fundamental issue contributing to fragmentation and poor coordination of healthcare delivery is that each member of the healthcare team, including patients, has a different view. Summarizing all of this care comprehensively for any member of the healthcare team is a “wickedly hard” visual analytics and data visualization problem to solve.

Keywords:

Data Display, Data Mining, Electronic Health Records

Introduction

The role of electronic health records (EHRs) has been described as both transformative and turbulent [1]. There are even arguments as to whether an EHR can be considered disruptive technology (in the traditional marketplace sense without fundamental change in the healthcare business model [2]), or is simply found to be disruptive to clinicians [3].

No longer just for documentation and communication about individual patients among a local healthcare team, EHRs are intended to support improvement in the overall quality and cost of healthcare [4,5]. Today's EHRs are expected to improve access to data across the continuum of care – irrespective of organizational boundaries [6]. Despite problems with clinical decision support components [7,8], EHRs are also expected to aid evidence-based care using experiential data [9], social determinants of health as related to a given patient [10], and personally supplied health data [11] – none of which have heretofore been included in individual health records. And, when data analytics are available to a clinician (separate from or as a product of an EHR), they must supply evidentiary information in real, or near real time [12].

Significantly less has been written about the environment of use and human factors associated with use – especially cognitive performance [13] and associated workflows. Indeed, a *Journal of the American Medical Association (JAMA)* Viewpoint in 2016 [14] suggests that “the evolution of EHRs has not kept pace with technology widely used to track, synthesize, and

visualize information in many other domains of modern life.”

There are a number of technologies that hold promise for improving the human experience in order to derive value from the EHR investment [15]. Some of these are older technologies, such as data display, registries, and registry functionality which typically have not been included in an EHR [16]. Some are newer, such as data/information visualization and visual analytics. The visual analytics discipline was initially founded in 2004 where the science of analytical reasoning with advanced interactive visual interfaces was challenged to analyze overwhelmingly disparate, conflicting, and dynamic information [17]. One of the earliest papers on visualization in healthcare for personal histories was published in 1996 [18].

In general, it is believed that the use of the term data visualization when applied in the literature to EHR systems is the result of a process of visual analytics. Data visualization may be static or interactive. When interactive, visual analytics supports the ability to act on the data in real time to glean additional perspectives. Interactive data visualization also puts control into the hands of the user, where details can be obtained ‘on demand’ [19]. Caban and Gotz [20] define visual analytics as “the science of analytical reasoning facilitated by advanced interactive visual interfaces.” These authors go on to state that “visual analytics techniques combine concepts from data mining, machine learning, human computing interaction, and human cognition.” This paper notes that “visual analytics systems – by combining advanced interactive visualization methods with statistical inference and correlation models – have the potential to support intuitive analysis for all ... user populations while masking the underlying complexity of the data.” [20] Finally, the statement made by Caban and Gotz that “the science of analytical reasoning facilitated by advanced interactive visual interfaces” makes the case for the authors to suggest that when a picture is worth a thousand words, data visualization is the picture created by a visual analytics technique applied to the thousand words [21]. The purpose of this paper is to provide a vision of data visualization's possibilities for leveraging new methodologies and techniques reflecting the diversity of health data and healthcare process.

Methods

A literature search was conducted using the terms data visualization and visual analytics focusing on real time or near real time EHR systems applied to clinical care. We also included the concepts of cognition or cognitive informatics. We avoided clinical decision support as too narrow, and description of tools without application to electronic health record data as too broad. The searches used PubMed. Due to the hybrid nature of visual analytics we also referred to a

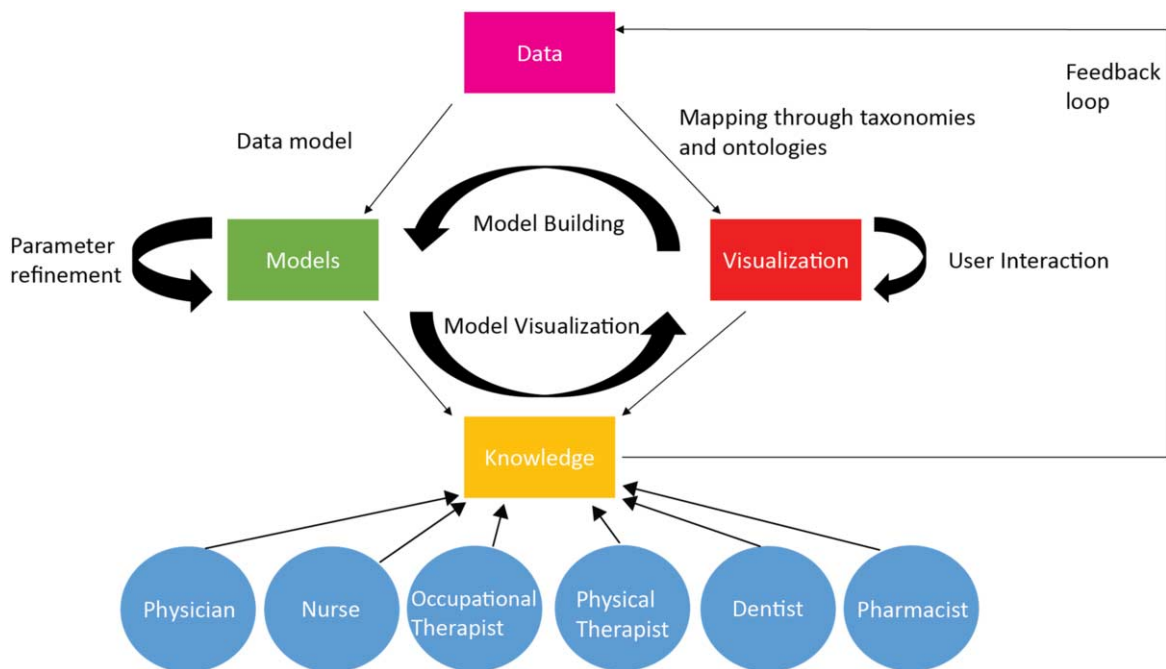


Figure 1 - Visual Data Exploration Inspired Across Multiple Health Disciplines adapted from Keim page 10 [19].

The bold contrasting colors between the data, models, knowledge, and visualization are designed to highlight the differences between the different areas of data exploration. The exploration begins with the data. The blue color of the health professional highlights the commonality of the biological underpinnings of all health professionals, however each circle is distinct, as professional training is different. Professional input into the figure is through the knowledge obtained from the models as the preexisting clinical health model will be different for each profession and the knowledge obtained will be different. The thicker lines represent manual steps throughout the process, where the thin lines can become automated.

number of text books and design books focused on visual learning.

Data Visualization in the Electronic Health Record

A total of 19 papers were identified that included use cases for data visualization of EHR data.

Table 1 - Papers in Data visualization

Author	Use case
Basole [22]	Explore care processes for complex pediatric asthma patients in Emergency Department (ED).
Dolan [23]	Interactive decision dashboard to help patients make decisions about unfamiliar healthcare management strategies.
Evans [24]	Automated case detection and response triggering system to monitor non-Intensive Care Unit (ICU) hospitalized patients every 5 minutes and identify early stages of physiologic deterioration.
Faiola [25]	Dashboard via EHR combines data from bedside monitoring devices, EHR, and medication pump display in ICU to improve data retrieval.
Farri [26]	Visual cues in EHR document user interface to support discovery of relevant patient information.
Foraker [27]	On demand data visualization tool concurrent with EHR navigation triggers Application

Hirsch [28] program interfaces (APIs) that collect parameters relative to cardiovascular health to render a risk profile for a given patient.

Huang [29] Data visualization tool displays patient information on a timeline and a problem cloud to facilitate review of essential patient information.

Huang [30] Visualization for temporal patterns of polymorbidities associated with complex chronic kidney disease (CKD) and its outcomes.

(Follow up study to [29]) Interactive web-based application enables researchers to study cohorts over time using EHR data; also can reveal potential trajectories of a disease over time.

Ledesma [31] Visualization library studied on various usability tests that enabled users to understand health data and its evolution over time.

Militello [32] Beta version of a screening and surveillance App (for colorectal cancer screening) used in the United States Veterans Health Administration's EHR.

Plaisant [33] Assessed the effectiveness of a data visualization tool to improve the display of medication lists to be reconciled (upon discharge).

Radhakrishnan [34] Use of visual analysis techniques to discover clinically salient associations between patient characteristics with problem-oriented health outcomes of older adult home health patients during the home health service period.

Ratwani [35]	Intuitive visualization dashboards for end users to facilitate exploration of patient safety event reporting systems and analyze trends.
Simpao [36]	An alert dashboard was created using override rates to provide rapid-cycle safety information.
Soulakis [37]	Visualization of collaborative EHR usage for hospitalized patients with heart failure was used to help strategically guide care coordination for patients at risk for readmission.
Warner [38]	A visualization tool specifically designed to be accessible to clinicians enables exploring for “patients like this one.”
Wongsupha sawat [39]	Summarization of temporal event data extracted from EHRs of a cohort of patients can be used to analyze disease (congestive heart failure) progression pathways and their outcomes.
Wongsupha sawat [40]	Study of the use of a data visualization tool to describe event sequences in patient transfers between (ED, ICU, intermediate care, and hospital ward) departments.

Predominately, and not surprisingly, these use cases focused on what typically are described as high risk clinical scenarios – disease conditions associated with many co-morbidities and likelihood of emergency department use and hospital readmission, or chronic and debilitating conditions that generally lead to complications [41]. While these areas of focus are important, they are atypical and limited in applicability to the overall processes of healthcare delivery. Examples of chronic conditions in need of models include pediatric asthma, chronic kidney disease, heart failure, diabetes, multiple sclerosis, and depression. In addition, predictors for physiologic deterioration and cardiovascular health issues were geared toward maintaining a stable condition. Intensive Care Unit (ICU) data analysis, transfers among care sites, and medication reconciliation are all challenging issues in busy hospitals and frequently the cause of hospital-acquired conditions. Several of the papers identified suggested that there was not only time savings in clinicians attempting to look for ‘patients like this one’ [38] to make more informed diagnosis and/or treatment plans, but efforts to learn from other patient experiences were actually made wherein the past the time required would often have kept clinicians from conducting such an exercise at all. Although not explicitly stated in any of the papers reviewed, collectively there appears to be the suggestion that interaction with data visualization – which can take from 2 to 5 minutes according to several of the papers – is acceptable for complex clinical cases.

While the focus of the literature review was on use of data visualization in association with an EHR, which implies a healthcare encounter (i.e., hospitalization, emergency department service, or office/clinic visit), patients are increasingly interested in becoming involved in their care. Being able to present information to help patients make treatment choices or lifestyle changes in a manner that is accessible for patients can significantly aid consumer engagement in their care. However, to engage this the latent meaning in the health data need to become accessible to non-experts in biomedical science, i.e., the lay perspective. Through cognitively accessible data, patients can engage in meaningful interactions with health experts to clarify and to improve more rapidly with the goal to readily reduce the likelihood of acquiring or exacerbating a condition [42].

Interactive data

The prior data visualization papers did not always include an explicit model, how the data is interrelated (Figure 1). Without models, data is just displayed and interaction is limited. The potential for various models to enable different filters to be applied based on the data needs of individual users or user types could allow a more interactive data presentation. It is well known that as many as a hundred different users may require access to a patient’s health record in a hospital. The use of such modeling to create truly interactive visualization techniques in order to understand the clinical condition of the patient from any given perspective appears yet to have been addressed.

Imagine, for instance, a nutritionist’s view of the patient’s needs as they dynamically change based on the narrowing of the physician’s differential diagnosis for the patient. One challenge unique to healthcare is that the knowledge base of professionals (physicians, nurses, social workers, pharmacists, physical therapists, occupational therapists, dentistry, nutritionist, public health and many more) is very broad. It appears that data visualization could be the means to aggregate and parse such massive knowledge bases by building models that support multivariate uses of data to improve the healthcare experience for patients.

Multivariate Datasets

Visualizing large continuous streams of structured data has been the focus to date, as described above. However, there are a number of other variables often not in such structured format. For example, lab values which arrive semiregularly can be updated as the results change from preliminary to final. Imaging modalities can vary from small size (amount of data) in data such as ultrasound, to the larger size in data including Magnetic Resonance Imaging, Computed tomography (CT), and many others. From all imaging modalities, the radiologist’s interpretation is normally provided in free text. A common viewing technique is either one large table, or in a timeline based approach. The challenge with both is the user is required to model the information mentally to arrive at knowledge. Creativity of clinicians and visual artists is the only limiting factor in creating solutions to these challenges.

Another area of data that needs better modeling and visualization, is the critical thinking of all health professionals. Without visualizing how professionals’ clinical judgement changes with new knowledge, continuously improving treatment protocols and impacting others’ knowledge cannot be achieved.

Visual Thinking

Visual thinking is a way to organize your thoughts and simplify complex concepts. In figure 1, we show how data can be incorporated into both models and visualization to enable visual thinking. Through better models, we better understand the refinement needed to further make better models and incorporate them into better visualizations. The data needs to be mapped through taxonomies and ontologies (invisible to the end user) in order to improve visualization within the EHR that can impact how clinicians can visually think about data to aid healthcare. Users’ interactions with visualization will create new knowledge. Just as knowledge is the application of experience to information, new knowledge will impact each professional based on personal experience. For example, an increase in a pain score (scale from 1 to 10) for a patient could indicate to a physician an additional medication may be needed,

where the nurse may recognize the level of pain as a failure in pain management education, or a physical therapist may identify the pain as the result of a therapy session to increase range of motion. Depending on when the patient's pain score is assessed and the timing of how the three professionals interact with the patient, all interpretations could be correct, only some could be correct, or none could be correct. Visualizing all of these interactions and building appropriate models across the range of disease states will be a challenge for the next decade. The creativity of the global community is required to solve hard challenges such as integrating multiple health professions into a single multiprofessional intuitive interface for something as common as a pain score and more innovation for the more complex health conditions encountered on a daily basis.

Designing new user interfaces for EHRs by applying these principles and new designs to engage clinicians both in the patient's working memory (part of short term memory that is concerned with immediate processing) and the health professional's long term memory is a huge challenge. Data visualization design and cognitive thought is not just about EHR usability criteria issued by United States National Institute of Standards and Technology (NIST). Rather, data visualization design and cognitive thought is about how healthcare professionals interact with the EHR. This paper focuses on a visual thinking process, and how the computer can engage both visual memory (a form of memory related to your visual experience) afforded through data visualization and working memory [43]. But even simple glyphs (icons), loaded with colors, shapes and direction, portray more information for working memory than just a list of data values. [43] For example, most EHR's list all of the laboratory values as simple numbers on a table. Applying these principals to ordinary laboratory results could improve time and comprehension of these data.

Discussion

The design and artistic thought and expression of models, visualization, and existing knowledge are not incorporated into the current EHRs. With better models of the clinical decision making of health professionals, the data becomes more amenable to visualizations. Through better measurement and representation, deeper knowledge about the patients and the work flow processes can be studied and improved. Through visual thinking research, refinement of the models, and collaboration with biomedical illustrators, a more intuitive and dynamic interface can enable clinicians to treat patients instead of spending hours reading computer screens.

Conclusions

The maturity of data visualization in healthcare is in its infancy. Visual analytics has been only cursorily applied to the healthcare field. Healthcare is one of the most complex of all disciplines, reflecting a diversity of trained professional team members in the care of an individual patient. For example, a single patient could encounter in one hospitalization, several different physicians, nurses, pharmacists, social workers, hospital administrators, physical therapists, occupational therapists, respiratory therapists, lab technicians, and many others. Each professional has a different view of the patient, and everyone contributes to the care of the patient differently. Through visual thinking and models new interactions with the EHR can improve speed of consuming information and memory. Arming these professionals and the patients themselves with visual thinking is a "wickedly hard" problem

to solve. As the fields of visual analytics and data visualization mature, the design criteria and improved usability will help improve the treatment of patients.

Acknowledgements

We would like to thank Mika Ishikawa, student in Health Information Management at the UIC, and Claire Heshmat, student in Biology at the UIC, for their assistance in conducting the literature search.

References

- [1] VL Patel, TG Kannampallil. Cognitive informatics in biomedicine and healthcare, *J Biomed Inform* 53 (2015), 3-14.
- [2] Wananmaker B, Bean D.. Why EHRs are not (yet) disruptive. 2013; Available at: <https://www.christenseninstitute.org/blog/why-ehrs-are-not-yet-disruptive/>. Accessed April 6th, 2017.
- [3] AK Jha, D Doolan, D Grandt, T Scott, DW Bates. The use of health information technology in seven nations, *Int J Med Inform* 77 (2008), 848-854.
- [4] A Sheikh, HS Sood, DW Bates. Leveraging health information technology to achieve the "triple aim" of healthcare reform, *J Am Med Inform Assoc* 22 (2015), 849-856.
- [5] JR Windle, TA Windle. Electronic Health Records and the Quest to Achieve the "Triple Aim", *J Am Coll Cardiol* 65 (2015), 1973-1975.
- [6] B Bell, K Thornton. From promise to reality: achieving the value of an EHR, *Health Financ Manage* 65 (2011), 50-56.
- [7] DL Hunt, RB Haynes, SE Hanna, K Smith. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review, *JAMA* 280 (1998), 1339-1346.
- [8] A Wright, TT Hickman, D McEvoy, S Aaron, A Ai, JM Andersen, et al. Analysis of clinical decision support system malfunctions: a case series and survey, *J Am Med Inform Assoc* 23 (2016), 1068-1076.
- [9] T Greenhalgh, J Howick, N Maskrey, Evidence Based Medicine Renaissance Group. Evidence based medicine: a movement in crisis? *BMJ* 348 (2014), g3725.
- [10] Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records, Board on Population Health and Public Health Practice, Institute of Medicine. (2014).
- [11] AE Chung, EM Basch. Incorporating the patient's voice into electronic health records through patient-reported outcomes as the "review of systems", *J Am Med Inform Assoc* 22 (2015), 914-916.
- [12] W Raghupathi, V Raghupathi. Big data analytics in healthcare: promise and potential, *Health Inf Sci Syst* 2 (2014), 3-2501-2-3. eCollection 2014.
- [13] RJ Holden. Cognitive performance-altering effects of electronic medical records: An application of the human factors paradigm for patient safety, *Cogn Technol Work* 13 (2011), 11-29.
- [14] DM Zulman, NH Shah, A Verghese. Evolutionary Pressures on the Electronic Health Record: Caring for Complexity, *JAMA* 316 (2016), 923-924.
- [15] L Riskin, R Koppel, D Riskin. Re-examining health IT policy: what will it take to derive value from our investment? *J Am Med Inform Assoc* 22 (2015), 459-464.
- [16] AS O'Malley, K Draper, R Gourevitch, DA Cross, SH Scholle. Electronic health records and support for primary care teamwork, *J Am Med Inform Assoc* 22 (2015), 426-434.
- [17] Thomas, J. J., Cook, Kristin A., National Visualization and Analytics Center., *Illuminating the path.* : IEEE Computer Society Los Alamitos, Calif., 2005.
- [18] LifeLines: visualizing personal histories. Proceedings of the SIGCHI conference on Human factors in computing systems: ACM; 1996.
- [19] D Keim, F Mansmann, J Schneidewind, J Thomas, H Ziegler. Visual analytics: Scope and challenges, *Visual data mining* (2008), 76-90.
- [20] JJ Caban, D Gotz. Visual analytics in healthcare--opportunities and research challenges, *J Am Med Inform Assoc* 22 (2015), 260-262.
- [21] Forum One Communications. Worth a Thousand Words: How to Display Health Data. 2014; Available at: <http://www.chcf.org/publications/2014/02/worth-thousand-words-data>. Accessed April/17, 2017.
- [22] RC Basole, ML Braunstein, V Kumar, H Park, M Kahng, DH Chau, et al. Understanding variations in pediatric asthma care processes in the emergency department using visual analytics, *J Am Med Inform Assoc* 22 (2015), 318-323.
- [23] JG Dolan, PJ Veazie, AJ Russ. Development and initial evaluation of a treatment decision dashboard, *BMC Med Inform Decis Mak* 13 (2013), 51-6947-13-51.
- [24] RS Evans, KG Kuttler, KJ Simpson, S Howe, PF Crossno, KV Johnson, et al. Automated detection of physiologic deterioration in hospitalized patients, *J Am Med Inform Assoc* 22 (2015), 350-360.

- [25] A Faiola, P Srinivas, J Duke. Supporting Clinical Cognition: A Human-Centered Approach to a Novel ICU Information Visualization Dashboard, *AMIA Annu Symp Proc* 2015 (2015), 560-569.
- [26] O Farri, A Rahman, KA Monsen, R Zhang, SV Pakhomov, DS Pieczkiewicz, et al. Impact of a prototype visualization tool for new information in EHR clinical documents, *Appl Clin Inform* 3 (2012), 404-418.
- [27] RE Foraker, B Kite, MM Kelley, AM Lai, C Roth, MA Lopetegui, et al. EHR-based Visualization Tool: Adoption Rates, Satisfaction, and Patient Outcomes, *EGEMS (Wash DC)* 3 (2015), 1159-9214.1159. eCollection 2015.
- [28] JS Hirsch, JS Tanenbaum, S Lipsky Gorman, C Liu, E Schmitz, D Hashorva, et al. HARVEST, a longitudinal patient record summarizer, *J Am Med Inform Assoc* 22 (2015), 263-274.
- [29] CW Huang, S Syed-Abdul, WS Jian, U Iqbal, PA Nguyen, P Lee, et al. A novel tool for visualizing chronic kidney disease associated polymorbidity: a 13-year cohort study in Taiwan, *J Am Med Inform Assoc* 22 (2015), 290-298.
- [30] CW Huang, R Lu, U Iqbal, SH Lin, PA Nguyen, HC Yang, et al. A richly interactive exploratory data analysis and visualization tool using electronic medical records, *BMC Med Inform Decis Mak* 15 (2015), 92-015-0218-7.
- [31] A Ledesma, M Al-Musawi, H Nieminen. Health figures: an open source JavaScript library for health data visualization, *BMC Med Inform Decis Mak* 16 (2016), 38-016-0275-6.
- [32] LG Militello, JJ Saleem, MR Borders, CE Sushereba, D Haverkamp, SP Wolf, et al. Designing Colorectal Cancer Screening Decision Support: A Cognitive Engineering Enterprise, *J Cogn Eng Decis Mak* 10 (2016), 74-90.
- [33] C Plaisant, J Wu, AZ Hettinger, S Powsner, B Shneiderman. Novel user interface design for medication reconciliation: an evaluation of Twinlist, *J Am Med Inform Assoc* 22 (2015), 340-349.
- [34] K Radhakrishnan, KA Monsen, SH Bae, W Zhang. Visual Analytics for Pattern Discovery in Home Care. Clinical Relevance for Quality Improvement, *Appl Clin Inform* 7 (2016), 711-730.
- [35] RM Ratwani, A Fong. 'Connecting the dots': leveraging visual analytics to make sense of patient safety event reports, *J Am Med Inform Assoc* 22 (2015), 312-317.
- [36] AF Simpao, LM Ahumada, BR Desai, CP Bonafide, JA Galvez, MA Rehman, et al. Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard, *J Am Med Inform Assoc* 22 (2015), 361-369.
- [37] ND Soulakis, MB Carson, YJ Lee, DH Schneider, CT Skeeahan, DM Scholtens. Visualizing collaborative electronic health record usage for hospitalized patients with heart failure, *J Am Med Inform Assoc* 22 (2015), 299-311.
- [38] JL Warner, JC Denny, DA Kreda, G Alterovitz. Seeing the forest through the trees: uncovering phenomic complexity through interactive network visualization, *J Am Med Inform Assoc* 22 (2015), 324-329.
- [39] Outflow: Visualizing patient flow by symptoms and outcome. IEEE VisWeek Workshop on Visual Analytics in Healthcare, Providence, Rhode Island, USA: American Medical Informatics Association; 2011.
- [40] LifeFlow: visualizing an overview of event sequences. Proceedings of the SIGCHI conference on human factors in computing systems: ACM; 2011.
- [41] S Howell, M Coory, J Martin, S Duckett. Using routine inpatient data to identify patients at risk of hospital readmission, *BMC Health Serv Res* 9 (2009), 96-6963-9-96.
- [42] J Laurance, S Henderson, PJ Howitt, M Matar, H Al Kuwari, S Edgman-Levitan, et al. Patient engagement: four case studies that highlight the potential for improved health outcomes and reduced costs, *Health Aff (Millwood)* 33 (2014), 1627-1634.
- [43] Ware C., Information visualization : perception for design. Morgan Kaufmann (Elsevier) Boston, MA, 2013.

Address for correspondence

Andrew D. Boyd (boyda@uic.edu) 1919 W Taylor St. MC 530
Chicago, IL 60612, USA. +1 (312) 996-8339.

What Information Does Your EHR Contain? Automatic Generation of a Clinical Metadata Warehouse (CMDW) to Support Identification and Data Access Within Distributed Clinical Research Networks

Philipp Bruland, Justin Doods, Michael Storck, and Martin Dugas

Institute of Medical Informatics, University of Münster, Münster, Germany

Abstract

Data dictionaries provide structural meta-information about data definitions in health information technology (HIT) systems. In this regard, reusing healthcare data for secondary purposes offers several advantages (e.g. reduced documentation times or increased data quality). Prerequisites for data reuse are its quality, availability and identical meaning of data. In diverse projects, research data warehouses serve as core components between heterogeneous clinical databases and various research applications. Given the complexity (high number of data elements) and dynamics (regular updates) of electronic health record (EHR) data structures, we propose a clinical metadata warehouse (CMDW) based on a metadata registry standard. Metadata of two large hospitals were automatically inserted into two CMDWs containing 16,230 forms and 310,519 data elements. Automatic updates of metadata are possible as well as semantic annotations. A CMDW allows metadata discovery, data quality assessment and similarity analyses. Common data models for distributed research networks can be established based on similarity analyses.

Keywords:

Information Systems; Metadata; Semantics

Introduction

Data dictionaries provide information about content, structure and in some cases, semantics of information systems. In the domain of health information technology (HIT), mostly pharmaceutical companies are equipped with such dictionaries to support the organization of their trial metadata, the creation process of case report forms (CRF) [1] and to semantically enrich CRFs' data elements with domain specific variable names based on e.g. the Study Data Tabulation Model (SDTM) [1] of the Clinical Data Interchange Standards Organization (CDISC). In the healthcare domain data dictionaries as well as metadata registries are also able to support the creation process of documentation structures within electronic health record (EHR) systems [2, 3].

A further use case of data dictionaries is the data discovery process within an information system infrastructure. This presupposes that metadata from relevant clinical IT systems is included and structured in a data dictionary. A catalog of available healthcare data is a great advantage to reusing routinely collected data for secondary purposes. For instance, clinical trial feasibility analyses, patient identification and recruitment or the execution of trial documentation can be

supported with data from EHR systems or data warehouses [4, 5, 6, 7, 8]. Furthermore, a prerequisite for the secondary use of healthcare data is its quality, the availability of desired elements and an identical meaning of already captured information within routine HIT systems [9].

Most of these projects underwent a laborious and time-consuming process of discovering desired data within clinical primary systems to prepare data exports. Since medical data are far away from being 100% structured, projects like SHARPN or cloud4health additionally focus on processing of free text clinical notes with natural language processing (NLP) techniques [10, 11].

Clinical research is oftentimes performed in a multi-center setting where data from several sites must be identified and analyzed. To achieve semantic interoperability and to identify matching data elements, it is essential that metadata of all participating systems are semantically annotated with codes of medical terminologies, [12] such as the International Classification of Diseases (ICD), the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) or the Unified Medical Language System (UMLS).

Data dictionaries or metadata registries can provide information about data elements and facilitate semantic annotation, which is usually not possible within clinical IT systems. However, populating such a metadata registry is mostly performed manually. Since EHR systems are built with a heterogeneous and proprietary database structure, it is impossible to have one schema for obtaining metadata information from several systems. On the other hand, such systems are not carved in stone, which means that data elements and templates are frequently modified. Given the vast amount of data elements (>1,000 per EHR system), it is almost impossible to manually administrate and synchronize the content of a data dictionary with reasonable efforts and in an appropriate period of time.

Data quality is a key issue for the secondary usage of medical data, especially for those data elements that are most frequently used in clinical trials [13]. Important aspects of data quality are origin, completeness and correctness [9], yet most data dictionaries do not contain any information about the quality of data which they are referencing to because a connection to the clinical IT systems is missing. Due to this fact, it is not possible to prepopulate (research) data warehouses or other applications with exported, predominantly structured patient data.

For several years, so-called distributed research networks are being established to conduct clinical research in cooperation

with different national or international institutions. Research network infrastructures such as PopMedNet, ODHSI or i2b2 SHRINE are based on a defined common data model or a core ontology to which clinical routine data is mapped [14, 15, 16]. Data queries can be performed based on a selection of these data schemata.

However, all of these research infrastructures require data that must be discovered and extracted from primary clinical IT systems to be populated. These tedious and time-consuming tasks of manual metadata processing and quality evaluation steps lead us to our study objectives. We aim to develop an automated generation approach for a clinical metadata warehouse (CMDW) based on HIT systems. The CMDW should serve as an intelligent metadata registry, which facilitates the identification, data quality assessment, and exporting process for the reuse of routinely collected medical data.

Methods

Diverse specifications for metadata models were analyzed such as: the ISO/IEC 11179-3:2013, openEHR, Clinical Element Models and the Operational Data Model (ODM) of CDISC [17, 18, 19, 20]. As a result, a generic metadata model was created to store system-specific information, templates (documentation forms), data elements, data quality measures, and relevant parameters to extract data from clinical IT databases. Additionally, permissible values for data elements and semantic annotations of all relevant entities within the CMDW were stored. MySQL was used as database and Java for the implementation of the CMDW.

Data structures in clinical IT systems are mostly organized heterogeneously. Therefore, a generic and universal approach for extracting meta-information from medical databases is essential in order to handle the diversity of systems. In this regard, we considered different schema matching approaches assembled by Rahm et al. [21]. For this step, we used Talend Open Studio – Data Integration [22]. This Extraction-Transform-Load (ETL) tool includes a comprehensive library of different database input and output connectors, which allows complex matching of definitions. Implementation in Talend was enhanced with a self-written Java code.

EHR systems of two major university hospitals (Agfa ORBIS from two sites and GE Healthcare Centricity Perinatal from One site) in western Germany were selected for the proof-of-concept of the CMDW. Database structures regarding templates and data elements were analyzed and a schema matching pipeline for the CMDW were developed.

Ethical approval was not required since only metadata parameters of EHR systems were processed to evaluate the software tool.

Results

The CMDW Data Model

We chose parts of the conceptual data model of the ISO/IEC 11179-3:2013, which specifies the structure of a metadata registry as basis for the database model of the clinical metadata warehouse. Adaptations were made in terms of the respective clinical IT system and are illustrated in Figure 1. Due to historical reasons, the documentation structure in most EHR systems is based on templates that are filled out for different purposes. Hence, a system contains one or more

templates, which again comprises several data elements. Departments are modeled in relationship to these templates to obtain additional information regarding the disease area of each template. Therefore, templates can be referenced to themselves in order to allow modelling of dependencies between closely related main- and sub-templates. According to the ISO 11179 standard, data elements may contain a value domain that includes a list of permissible values (value sets) as options from which to choose. A data element and value domain are then linked in the common circle with the data element concept and conceptual domain classes. To address the assignment of semantic codes, a link between code-system-codes to templates, data elements, value domain and permissible values is established. Lastly, codes are assigned to a specific code system.

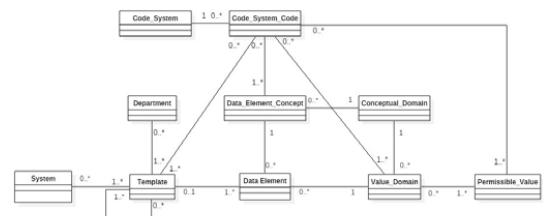


Figure 1 – Entity Relationship Model of the Clinical Metadata Warehouse referring to the ISO 11179:3 Model.

Metadata Integration Process

Metadata is integrated into the CMDW by the use of ETL pipelines within Talend Open Studio. For all three EHR systems an Oracle database connector was used. Figure 2. shows the pipeline for the metadata integration step.

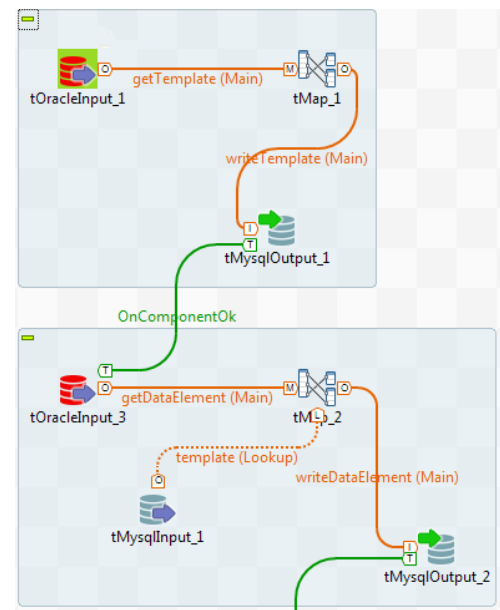


Figure 2 –Extraction of the ETL Pipeline between Templates and Data Elements to Integrate EHR Metadata into the CMDW.

First, all templates from a given system were selected with name, description, type (main-template, administrative-

template, sub-template associated with other templates, etc.), usage status, version, and system internal ID. Next, data elements were extracted and saved with their relationship to the templates. Again, internal ID, name, description and data type were stored. Then a link to the medical departments was created and permissible values were extracted. Semantic codes were not assigned due to missing annotations in the source systems. This process needs to be performed manually or semi-automatically by eligible programs.

This ETL pipeline can be modified with respect to system specific metadata schemata of all relevant entities. Using Talend Open Studio, this process took approximately 14 seconds to automatically insert all metadata information from one EHR system. The same pipeline can be applied to update the metadata directly, taking several minutes to check for new, deleted or modified entities in our proof-of-concept setting.

Metadata from three EHR systems were integrated into the CMDW. The first EHR system contained 8,135 templates which were referenced to 37 departments and 155,480 data elements. A smaller quality assurance system contained 83 templates and a catalog including 6,102 available and unique data elements. The EHR system of the second site comprised 8,012 templates and 148,937 data elements. All of these were attached with system internal IDs to enable patient data exports into a data warehouse for further analyses and to facilitate an automatic calculation of data quality. To assess completeness for a patient data element, the amount of filled data elements was divided by the total number of patients for a defined time period (e.g. current year).

Overall System Architecture

To not only comply with local and national data protection laws, but also with the issue of data sovereignty, we decided to choose a distributed decentralized approach for the architecture (see Figure 3).

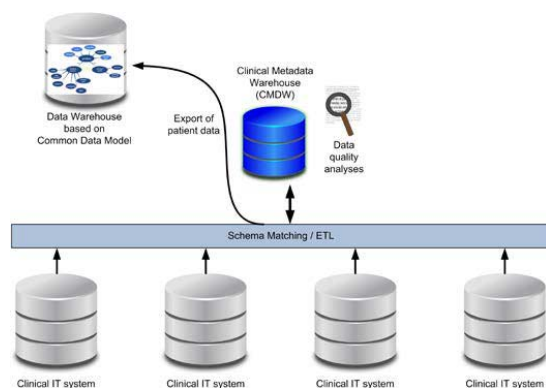


Figure 3 – Architecture of the Clinical Metadata Warehouse approach.

Metadata from clinical IT systems was matched towards the site central CMDW database and transferred. The CMDW is a repository that contains cross-references and meta information from different IT systems. CMDW allowed browsing through the data inventory of all elements with the option to filter for specific medical departments. Internal database IDs of each EHR system were stored in the CMDW to perform data quality analyses and patient data exports into a common data model, such as OMOP (from OHDSI), PCORnet (from PopMedNet), or a clinical or research data warehouse

infrastructure. Accessing patient data in a centralized architecture would not be feasible due to data protection laws.

Discussion

The Clinical Metadata Warehouse enables automatic generation of data inventory, which is populated with metadata of any primary clinical IT system. It substitutes the previously tedious, time-consuming and manual process of establishing such an inventory. Metadata within the CMDW can be updated automatically, which is a major advantage due to the frequent modifications of clinical IT systems. Common metadata registries or data dictionary solutions often contain only a limited predefined set of data elements for a specific purpose. With a clinical metadata warehouse, not only can selected data elements be considered, but also the full set of structural data contained in HIT systems.

Schema matching is performed with Talend Open Studio [21], which is a flexible and powerful ETL tool for processing and manipulating data between arbitrary sources and targets. For the metadata integration into the CMDW, other ETL tools or schema matching approaches are also feasible. A compilation and evaluation of further toolkits that support matching of schemata is described by Bellahsene et al. [23].

Sun et al. are following a similar approach using ontologies and semantic web technologies [24]. Their Semantic Data Virtualization solution is also an intelligent layer between non-semantically enriched EHR source systems and data consumers. In this layer, source data is mapped to RDF (Resource Description Framework) data with semantics. They mention the burden of common data warehouse approaches to keep the data synchronized between the data source and the target warehouse [24]. Similar to this work, Mate et al. also used an ontological description layer for their EHR data and they are also able to automatically generate SQL queries for data extraction [25]. However, all of these promising approaches suffer from the drawback of manual association and extraction of metadata.

Overall, metadata registries also support the creation process of CRFs in clinical trials [26, 27]. Harmonized data elements and documentation templates could be reused to create comparable results and to reduce the time and efforts for developing clinical trial documentation. Pharmaceutical companies such as Amgen, Lilly, Novartis or Roche maintain a metadata registry with semantically annotated data elements. These elements are mostly annotated with codes of the SDTM of CDISC, which are used for submission of clinical trial data to regulatory authorities.

Data dictionaries or metadata registries have a substantial effect on data sharing across clinical systems, which is also supported by Hicken et al. [28]. In this regard, clinical research networks such as PopMedNet, ODHSI or i2b2 SHRINE facilitate the definition of common data models or core ontologies to define data points that could be queried via a comprehensive network of institutions. The CMDW approach is able to support such research infrastructures by identifying relevant data points in EHR systems and fostering the data population process of data warehouses or other research infrastructures.

Strengths and Weaknesses

The ISO 11179 standard was chosen for the CMDW because it was effective in related work [29]. In addition, it is widely used in diverse projects such as the National Cancer Institute's

caDSR (Cancer Data Standards Registry and Repository) [2], Australia's METeOR (Metadata Online Registry) [3], and USHIK [4].

Currently, both CMDWs (one for each site) are only populated with data from three EHR systems. Extracting data from EHRs for quality evaluation or secondary purposes greatly relies on the underlying structure of the data model. It is likely that other systems have a different database structure, making the automatically generated SQL queries for data extraction inoperable. Therefore, we are working on a general approach for mapping metadata with clinical data tables.

Further Research

Since most EHR systems do not provide semantically annotated data elements, the annotation task could be supported by the CMDW in the following ways: (1) manually adding semantic codes for a corresponding data element; (2) implementing a CDISC ODM export for certain forms to support reusing existing elements and semantic annotation; (3) developing an ODM importer to update corresponding elements; and (4) integrating semantic annotation tools, such as cTAKES [5] or MetaMap [6] to support the coding process of metadata elements.

Additional attention should focus on the quality of the schema matchings. Different "automatic" schema matching approaches exist [21], which could support the process developing guidelines and testing mechanisms to evaluate the mapping process of EHR databases.

Lastly, the identification of redundant routine data should be supported based on metadata content of all systems available within the CMDW. This could be promoted by tools and methodologies that are implemented in ODMSummary [30]. Semantic codes, names and permissible values are used for the comparison to decide whether two data elements are identical, matching, and transformable, etc. These results could be used to improve routine clinical documentation by harmonizing documentation structures and reusing previously entered patient data.

Further research will emphasize the standardized connection between data warehouses and different research infrastructures to evaluate the practicability of this approach.

Conclusion

The automatic generation and update of a clinical metadata warehouse based on EHR systems is feasible and supports the identification and evaluation of potentially relevant patient data for secondary purposes. Given the large number of data elements in EHR systems, this approach is advantageous to metadata registry approaches with a limited manual set of data elements.

Acknowledgement

Funded by German Ministry of Research (BMBF Grant 01ZZ16028)

References

- [1] Dugas M. Design of case report forms based on a public metadata registry: re-use of data elements to improve compatibility of data. *Trials* **17** (2016), 566.
- [2] Nadkarni PM, Brandt CM, Marenco L. WebEAV: Automatic metadata-driven generation of web interfaces to entity-attribute-value databases. *J Am Med Inform Assoc* **7** (2000), 343-56.
- [3] Bürkle T, Prokosch HU, Michel A, Dudeck J. Data dictionaries at Giesen University Hospital: past-present-future. *Proc AMIA Symp.* 1998:875-9.
- [4] Bruland P, Forster C, Breil B, Ständer S, Dugas M, Fritz F. Does single-source create an added value? Evaluating the impact of introducing x4T into the clinical routine on workflow modifications, data quality and cost-benefit. *Int J Med Inform* **83** (2014), 915-28.
- [5] Dziuballe P, Forster C, Breil B, Thiemann V, Fritz F, Lechtenbörger J, Vossen G, Dugas M. The single source architecture x4T to connect medical documentation and clinical research. *Stud Health Technol Inform* **169** (2011), 902-906.
- [6] De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claehtout B, Karakoyun T, Ohmann C, Lastic PY, Ammour N, Kush R, Dupont D, Cuggia M, Daniel C, Thienpont G, Coorevits P. Using electronic health records for clinical research: the case of the EHR4CR project. *J Biomed Inform* **53** (2015), 162-73.
- [7] Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, Shirey-Rice J, Kirby J, Harris PA. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* **52** (2014), 28-35.
- [8] Hersh WR. Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Am J Manag Care* **13** (2007), 277-278.
- [9] Weiskopf NG, Hripesak G, Swaminathan S, Weng C. Defining and measuring completeness of electronic health records for secondary use. *J Biomed Inform* **46** (2013), 830-836.
- [10] Pathak J, Bailey KR, Beebe CE, Bethard S, Carrell DC, Chen PJ, Dligach D, Endle CM, Hart LA, Haug PJ, Huff SM, Kaggal VC, Li D, Liu H, Marchant K, Masanz J, Miller T, Oniki TA, Palmer M, Peterson KJ, Rea S, Savova GK, Stanel CR, Sohn S, Solbrig HR, Suesse DB, Tao C, Taylor DP, Westberg L, Wu S, Zhuo N, Chute CG. Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPN consortium. *J Am Med Inform Assoc* **20** (2013), e341-e348.
- [11] Christoph J, Griebel L, Leb I, Engel I, Köpcke F, Toddenroth D, Prokosch HU, Laufer J, Marquardt K, Sedlmayr M. Secure Secondary Use of Clinical Data with Cloud-based NLP Services. Towards a Highly Scalable Research Infrastructure. *Methods Inf Med* **54** (2015), 276-82.
- [12] Dugas M. Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems. *Methods Inf Med* **53** (2014), 516-517.
- [13] Bruland P, McGilchrist M, Zapletal E, Acosta D, Proeve J, Askin S, Ganslandt T, Doods J, Dugas M. Common data elements for secondary use of electronic health record data for clinical trial execution and serious adverse event reporting. *BMC Med Res Methodol* **22** (2016), 159.
- [14] PopMedNet™. Distributed health data network. <http://www.popmednet.org>. Accessed: December 2016.
- [15] Hripesak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform* **216** (2015), 574-578.
- [16] Weber GM, Murphy SN, McMurphy AJ, Macfadden D, Nigrin DJ, Churchill S, Kohane IS. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* **16** (2009), 624-30.
- [17] ISO/IEC 11179-3:2013, Information Technology – Metadata Registries (MDR) - Part 3: Registry metamodel and basic attributes.
- [18] Kalra D, Beale T, Heard S. The openEHR Foundation. *Stud Health Technol Inform* **115** (2005), 153-73.
- [19] Coyle JF, Mori AR, Huff SM. Standards for detailed clinical models as the basis for medical data exchange and decision support. *Int J Med Inform* **69** (2003), 157-174.
- [20] CDISC Operational Data Model (ODM). <http://www.cdisc.org/odm> Accessed: December 2016
- [21] Rahm E, Bernstein PA. A Survey of Approaches to Automatic Schema Matching. *The VLDB Journal* **10** (2001), 334-350.
- [22] Talend Open Studio for Data Integration. <http://www.talend.com/products/data-integration> Accessed: December 2016.
- [23] Bellahsene Z, Bonifati A, Rahm E. Schema Matching and Mapping. Springer, 2011. 33.
- [24] Sun H, Depraetere K, De Roo J, Mels G, De Vloed B, Twagirumukiza M, Colaert D. Semantic processing of EHR data for clinical research. *J Biomed Inform* **58** (2015), 247-259.

- [25] Mate S, Köpcke F, Toddenroth D, Martin M, Prokosch HU, Bürkle T, Ganslandt T. Ontology-based data integration between clinical and research systems. *PLoS One* **10** (2015) e0116656.
- [26] Warzel DB, Andonaydis C, McCurry B, Chilukuri R, Ishmukhamedov S, Covitz P. [Common data element \(CDE\) management and deployment in clinical trials](#). AMIA Annu Symp Proc. 2003:1048.
- [27] Merzweiler A, Knaup P, Weber R, Ehlerding H, Haux R, Wiedemann T. Recording clinical data--from a general set of record items to case report forms (CRF) for clinics. *Stud Health Technol Inform* **84** (2001), 653-7.
- [28] Hicken VN, Thornton SN, Rocha RA. Integration challenges of clinical information systems developed without a shared data dictionary. *Stud Health Technol Inform* **107** (2004), 1053-1057.
- [29] Bruland P, Breil B, Fritz F, Dugas M. Interoperability in clinical research: from metadata registries to semantically annotated CDISC ODM. *Stud Health Technol Inform* **180** (2012), 564-568.
- [30] Storck M, Krumm R, Dugas M. ODMSummary: A Tool for Automatic Structured Comparison of Multiple Medical Forms Based on Semantic Annotation with the Unified Medical Language System. *PLoS One*. **11** (2016), e0164569.

Address for correspondence

Philipp Bruland, PhD
 Institute of Medical Informatics, University Hospital Münster
 Albert-Schweitzer-Campus 1 – Building A11
 48149 Münster, Germany
philipp.bruland@uni-muenster.de

Automatic Identification of Glaucoma Using Deep Learning Methods

Allan Cerentini^a, Daniel Welfer^a, Marcos Cordeiro d'Ornellas^a,
Carlos Jesus Pereira Haygert^b, Gustavo Nogara Dotto^b

^aGraduate Program in Computer Science (PPGI), Department of Applied Computing (DCOM),
Santa Maria, Rio Grande do Sul, Brazil

^bDepartment of Clinical Medicine, Santa Maria University Hospital,
Federal University of Santa Maria (UFSM), Rio Grande do Sul, Brazil

Abstract

This paper proposes an automatic classification method to detect glaucoma in fundus images. The method is based on training a neural network using public image databases. The network used in this paper is the GoogLeNet, adapted for this proposal. The methodology was divided into two stages, namely: (1) detection of the region of interest (ROI); (2) image classification. We first used a sliding-window approach combined with the GoogLeNet network. This network was trained using manually extracted ROIs and other fundus image structures. Afterwards, another GoogLeNet model was trained using the previous resulting images. Then those images were used to train another GoogLeNet model to automatically detect glaucoma. To prevent overfitting, data augmentation techniques were used on smaller databases. The results demonstrated that the network had a good accuracy, even with poor quality images found in some databases or generated by the data augmentation algorithm.

Keywords:

Glaucoma; Retina; Neural Network (Computer)

Introduction

Glaucoma is a lesion that occurs inside of the optic nerve, which can cause low vision or even total blindness. Currently, this lesion has no cure, but an early diagnosis with treatment can prevent the disease progression [1] [2]. One of the approaches used to detect disease is called ophthalmoscopy. This examination may generate a fundus image, which is used for this work. Currently, the detection of glaucoma is done manually by retina experts. They use the ratio between optic nerve size and lesion size, to diagnose the presence and severity. It is a process that takes time, and the lesion is not always clear. This work proposes to perform this task automatically. To perform this task, we have to deal with some challenges, namely: very subtle lesions, poor image quality, and illumination problems. There are related works that extract the size of the optic nerve and the lesion to perform this classification; others use machine-learning methods as neural networks for this purpose. This work aims to apply a robust neural network, developed to deal with hard image classification challenges, in order to overcome many of the problems presented previously. This paper presents a summary of some similar works, description of images datasets used, methodology applied, and the results. For automatic classification, two methods were developed: one to find the region of interest (ROI), optic nerve, in the fundus image and another to classify this region between healthy and with glaucoma.

Chen et al. [3] described a method that uses a neural network with four hidden layers. The input layer accepts images with a dimension of 256x256 and three color channels. The datasets used were artificially augmented using random cropping techniques and mirroring. The work performed two experiments using two image bases, (1) The ORIGA-light, which has 650 images and, (2) SCES, which has 1676 images. The first experiment used 99 ORIGA-light images to train the network and 551 for test. The second used all the ORIGA images to train the network and all of the SCES for test. The accuracy of both experiments was 83.12% and 88.7% respectively.

The method described by Sheeba et al. [4] also uses a neural network, but with two hidden layers. There is no clear information about the private dataset provided by Giridhar Eye Institute in Cochin. It consists of 20 images with unknown ground truth, where 28 have glaucoma and 12 are normal. These were processed using erosion and dilation techniques, where the dilated image is used as background and subtracted from the grayscale image obtained. The resulting intensity is adjusted, and then it is converted to binary values by thresholding the optical disc region. The network was trained by using 20 images with unknown label, and tested in the 40 remaining images. They reported that 34 images were classified correctly.

Zhang et al. [5] consists of an online repository of fundus images. The goal is to provide a way to share those images with the public. Researchers also can benchmark their algorithms. A segmentation and classification tool was developed to assist in the construction of the image database. At the date the article was published, the database was composed of 650 images, delimited by specialists. The ratio between the size of the excavation and the size of the lesion was used for classification. The detection of the region of interest was performed automatically. The first stage is the preprocessing of images, like fringe remove. The fringe happens due to patients not putting their eye so close to the capture machine, generating a light entrance by the edges. This light input hinders the ROI's detection, since it is done using the centroid of the brightest part of the image. This method has 96% accuracy according to the article. In cases where ROI is not found by the system, the user can manually select that region.

Considering the previous works, we noticed that literature already presents works using neural networks, to detect the presence of glaucoma. However, the results with good accuracy are based on good quality images. The purpose of our work is to perform this detection even in images with low contrast, high amount of noise and low resolution. And, to detecting the ROI of those images using a robust neural network.

Materials and Methods

Image Databases

We used four public image databases that were used for training and a validation of the network. The first database, High Resolution Backgrounds (HRF) [6], was composed of 45 fundus images, divided into 3 groups of 15 images. A group of images had glaucoma, while the other groups had healthy eyes and, the last, diabetic retinopathy. For the test and validation, only the groups with glaucoma and normal images were selected.

The remaining three databases, RIM-ONE r1, RIM-ONE r2 and RIM-ONE r3, were part of the work done by Fumero *et al.* [7]. RIM-ONE r1 and RIM-ONE r2 were composed of ROI images and RIM-ONE r3 of stereo fundus images. The first database had 40 images with some degree of glaucoma and 118 healthy images, while the second had 200 images with glaucoma and 225 healthy images. The third had 74 images with glaucoma and 85 healthy images. Each image has a ground truth delimited by a specialist. For the third database, the stereo images were divided in two parts and only the diagnosed part was used in the experiment.

Deep Learning

Artificial neural networks are part of a set of techniques in the Machine Learning area. This technique was inspired by the learning process of a biological brain, and is constructed using a fully connected neural network. Briefly, each neuron receives input signals from several sources, as well as from other neurons. Each entry is multiplied by a value, called weight. All results are then summed, and verified in an activation function that decides whether the neuron should send a signal ahead. The learning occurs by adjusting those weights. When a set of inputs gives a wrong output value, the weights are adjusted to make that output correct. Trying not to disturb the result of another set of inputs [8]. The depth of an artificial neural network is measured according to the number of layers between the input and the output of a network, they are called hidden layers. An input signal passes through the network until the last layer, called output, is reached. The last layer gives the prediction of that input signal. A larger number of hidden layers allow for the classification of more complex data, such as images [9].

Deep learning is the process of training multi-layered neural networks. To facilitate this process, we will use an open source framework developed by Google called Tensorflow [10]. It is the second generation of a large-scale machine learning system developed by Google. The system originated from a project called Google Brain started in 2011, where they built the DistBelief, which was the first generation. Tensorflow uses tensors, a multi-dimensional array, to represent data as images. Graphs are employed to represent the flow of operations, where each node represents an operation with zero or more tensors. This framework has implementations of several types of algorithms, mathematical models and specific functions optimized for the training of neural networks. It also provides a practical way to use a graphics processing unit (GPU) to accelerate the training process.

An existing neural network model, called GoogLeNet[11], was used. This model was developed by Google and competed in the ImageNet challenge from 2014, whose objective was to classify about 1.3 million of images in one of the 1000 different classes. The goal of Google was to create a computationally efficient network that could be run even with low computation resources. Some of techniques employed like batch normalization, residual connection and factorization allowed for

increased accuracy while maintaining performance. The network can be acquired previously trained, with the pre-defined weights. For this work the network was modified to train only a number of classes needed to solve some problems. The weights were maintained to use the principle of learning transfer [12].

Region of Interest (ROI) Detection

Finding manually the ROI of each image, from a large image database, is a hard process. An algorithm was developed to automatically detect this region. This is based on object detection works using neural networks from Malisiewicz *et al.* [13], Sermanet *et al.* [14] and an ROI detection work from Xu *et al.* [15]. A neural network was trained using the GoogLeNet model to classify images into two categories: region of interest or background. For this process, 107 images from ROI and 4693 from other fundus regions were taken from HRF database. An algorithm was made to make those crops, and they were manually classified. It was not a difficult task, because the crops were sequential. The network was trained and performed at about 99% accuracy, distinguishing ROI images from other structures from fundus. This good detection accuracy is due to the distinct characteristics presented in the images.

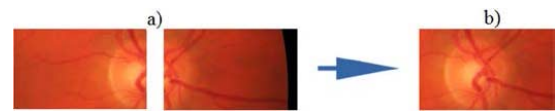


Figure 1 – The first step of the algorithm is to verify which are the possible ROI in the fundus image, resulting in a) Overlapping sliding windows. Then a suppression method is applied, resulting in only one location b) Result after suppression algorithm

To find the region of interest within the image, an algorithm was developed that uses sliding windows to scan the image in search of that region. Each window has size proportional to the size of the image. For each window it is checked on the previously described network if it is over the region of interest. If it is then the coordinates of this window are saved. If there is more than one case where this happens then a suppression method based on Malisiewicz *et al.* [13] and Xu *et al.* [15] is used, to avoid cases like that of Figure 1, where each window detects a part of the region of interest. This leads to loss of accuracy, since we need the whole region for classification.

Data Augmentation Process

The data augmentation process consists of adding deformations and noise to a data set, with the purpose of increasing the amount of data available for training. Providing a better accuracy to predict data from other datasets [16]. This technique is often used in small datasets, in order to prevent the network from eventually learning characteristics that are not relevant to classification, such as: bright images being classified as glaucoma and low contrast images as normal. It is a process that also prevents overfitting, which is a process where the network begins to memorize the images instead of learning about them. This also causes the network to have a good accuracy in the trained dataset and bad accuracy in others. The work proposed by Wu *et al.* [17] demonstrates the efficacy of this process to eliminate such problems presented previously, where several transformations in images are used for the learning processes focus on key features in the images.

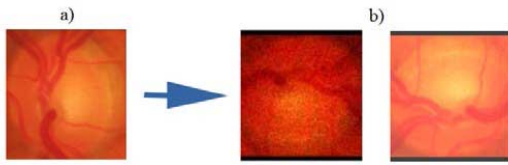


Figure 2 – Image resulting from ROI extraction algorithm
a) Extracted ROI from HRF database. Due to the small amount of images available in this database, transformations were applied. b) Results from data augmentation process.

In this work, a rotation of 90, 180 or 270 degrees is randomly applied on each image, as well size rescaling, gamma variation and addition of Gaussian noise. We can apply this process several times in an image. For each iteration we add an image to the training database. Figure 2 demonstrates an example of an image being transformed in two extra images for the training.

Image Classification

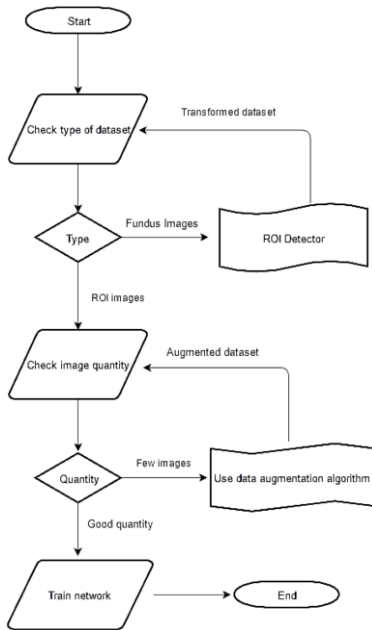


Figure 3 – Dataset training workflow

For the glaucoma classification, a neural network also based on the GoogLeNet, was trained. The training process is presented in Figure 3 and was applied to all image databases. First, we have to verify if it is a fundus or ROI database. The images must include the ROI for the classification. If it has fundus images, then all its images go through the process of ROI detection, described earlier. This process transforms a set of fundus images into a set of images of ROI images. Then we can verify if this dataset contains the minimum amount of images, to avoid problems like overfitting. Databases with less than 100 images per category were classified as "few images". In this case the image augmentation algorithm is used to expand the quantity of this category into an arbitrary value greater than the minimum value. When the quantity is good, the database is trained in the network. The processing was applied to HRF, Rim-one-r1 and Rim-one-r3 databases. The HRF database went through the data augmentation process,

turning its 30 images into 330. The Rim-one-r1 database glaucoma imaging class also went through this process, the number of images increased from 40 to 120. The Rim-one-r3 database went through ROI detection process.

Experimental results

The results obtained in this experiment are shown in Table 1. They were obtained by using 10% of dataset images to validate the accuracy of network. These images were randomly selected and were not used for neural network training. This helps us to evaluate a network more precisely, by simulating images from other image databases. The training process used 5000 training steps, the number of times a set of images from the data set passes through the network. And with a learning rate, sensitivity adjustment in weights, at 0.01 per update. Training each dataset took about 7 minutes using a GeForce Gtx 1070. For the final test, involving all the image databases, the artificially generated images, generated by the data augmentation algorithm were removed.

Table 1 – Success rate in the glaucoma detection using the proposed method

Image Database	Accuracy
HRF	90,0%
RIM-ONE r1	94,2 %
RIM-ONE r2	86,2 %
RIM-ONE r3	86,4 %
HRF + RIM-ONE r1 + RIM-ONE r2 + RIM-ONE r3	87,6%

We obtained a result close to the experiment described in the work done by Chen et al. [3], but with different databases. It is a satisfactory result, considering the amount of images considered difficult for this classification, such as: large variation in brightness, much noise and ROI barely visible. Some images are considered problematic are presented in Figure 4. There are no post-processing techniques to improve the quality of images. As well neither the common practice of removing blood vessels from those images.



Figure 4 – Examples of challenging images, due to low contrast or much noise.

Conclusion

In this paper a deep learning method was used, to detect the presence of glaucoma in the fundus images. The GoogLeNet neural network model from Google was used to accomplish this classification. It was able to detect the presence of glaucoma even in images where it appears only subtly, and in images with very low quality, generated by data augmentation. We obtained 90% accuracy on HRF database, RIM-ONE r1 with 94,2% accuracy, RIM-ONE r2 with 86,2% accuracy, and RIM-ONE r3 with 86,4% accuracy. The combination of all

databases resulted in 87,6% accuracy. To increase the amount of data of some networks, helping the training, a data augmentation algorithm was used. For the extraction of ROI from fundus image databases an algorithm was developed to detect this area. For future work a pre-processing can be done on the images of the databases, as well as expanding the number of databases of images for testing.

References

- [1] D. B. Henson and R. Thampy, "Preventing blindness from glaucoma: Better screening with existing tests should be the priority," *BMJ Br. Med. J.*, vol. 331, no. 7509, pp. 120–121, Jul. 2005.
- [2] L. Rossetti et al., "Blindness and Glaucoma: A Multicenter Data Review from 7 Academic Eye Clinics," *PLoS One*, vol. 10, no. 8, p. e0136632, Aug. 2015.
- [3] X. Chen, Y. Xu, D. W. Kee Wong, T. Y. Wong and J. Liu, "Glaucoma detection based on deep convolutional neural network," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 715-718.
- [4] O. Sheeba, J. George, P. K. Rajin, T. Nisha, G. Sherin, "Glaucoma Detection Using Artificial Neural Network", *International Journal of Engineering and Technology*, vol. 6, no. 2, pp. 158-161, 2014.
- [5] Z. Zhang et al., "ORIGA-light: An online retinal fundus image database for glaucoma analysis and research," 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, 2010, pp. 3065-3068.
- [6] J. Odstrcilik, R. Kolar, T. Kubena, P. Cernosek, A. Budai, J. Hornegger, J. Gazarek, O. Svoboda, J. Jan and E. Angelopoulou, "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database", *IET Image Processing*, vol. 7, no. 4, pp. 373-383, 2013.
- [7] F. Fumero, S. Alayon, J. Sanchez, J. Sigut and M. Gonzalez-Hernandez, "RIM-ONE: An open retinal image database for optic nerve evaluation", 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), 2011.
- [8] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85-117, 2015.
- [9] Y. Bengio, A. Courville and P. Vincent, "Representation Learning: A Review and New Perspectives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, 2013.
- [10] M. Abadi et al., "TensorFlow: A System for Large-Scale Machine Learning," *USENIX Symp. Oper. Syst. Des. Implement.* vol. abs/1605.08695, 2016.
- [11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv:1512.00567*, 2015.
- [12] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Adv. Neural Inf. Process. Syst.* 27 (Proceedings NIPS), vol. 27, pp. 1–9, 2014.
- [13] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 89–96.
- [14] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," *arXiv Prepr. arXiv*, p. 1312.6229, 2013.
- [15] Y. Xu, D. Xu, S. Lin, J. Liu, J. Cheng, C. Cheung, T. Aung and T. Wong, "Sliding Window and Regression Based Cup Detection in Digital Fundus Images for Glaucoma Diagnosis", *Lecture Notes in Computer Science*, pp. 1-8, 2011.
- [16] J. Salamon and J. P. Bello, "Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification," *arXiv Prepr. ArXiv*1608.04363, 2016.
- [17] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, "Deep Image: Scaling up Image Recognition," *Arxiv*, p. 12, 2015.

Addresses for correspondence

Allan Cerentini, acerentini@inf.ufsm

Daniel Welfer, welfer@gmail.com

Marcos Cordeiro d'Omeillas, marcosdomellas@gmail.com

Mining Adverse Drug Reactions in Social Media with Named Entity Recognition and Semantic Methods

Xiaoyi Chen^a, Myrtille Deldossi^b, Rim Aboukhamis^c, Carole Faviez^d, Badisse Dahamna^{e,f,g}, Pierre Karapetiantz^a, Armelle Guenegou-Arnoux^a, Yannick Girardeau^{h,i}, Sylvie Guillemain-Lanne^b, Agnès Lillo-Le-Louët^c, Nathalie Texier^d, Anita Burgun^{a,h,i}, Sandrine Katsahian^{a,h,i}

^a INSERM, UMRS1138, équipe 22, Centre de Recherche des Cordeliers, Paris, France

^b Expert System, 75012 Paris, France

^c Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France

^d Kappa Santé, 75002 Paris, France

^e Service d'Informatique Biomédicale, CHU de Rouen, France

^f LITIS-TIBS EA 4108, 76031 Rouen Cedex, France

^g INSERM, U1142, LIMICS, 75006 Paris, France

^h Université Paris Descartes, Sorbonne Paris Cité, UMRS1138, Centre de Recherche des Cordeliers, Paris, France

ⁱ Département d'Informatique Hospitalière, Hôpital Européen Georges-Pompidou, AP-HP, Paris, France

Abstract

Suspected adverse drug reactions (ADR) reported by patients through social media can be a complementary source to current pharmacovigilance systems. However, the performance of text mining tools applied to social media text data to discover ADRs needs to be evaluated. In this paper, we introduce the approach developed to mine ADR from French social media. A protocol of evaluation is highlighted, which includes a detailed sample size determination and evaluation corpus constitution. Our text mining approach provided very encouraging preliminary results with F-measures of 0.94 and 0.81 for recognition of drugs and symptoms respectively, and with F-measure of 0.70 for ADR detection. Therefore, this approach is promising for downstream pharmacovigilance analysis.

Keywords:

Social Media; Pharmacovigilance; Data Mining

Introduction

The rapid expansion of the Internet and social media is changing the way people gather information about disease and treatment, as well as how they share personal health experiences with others [1]. The *Digit in 2016* [2] reported that, in France, 86% of the population are active internet users. This proportion is higher than Western Europe's average (83%) and slightly lower than North America's average (88%). Various questionnaire statistics [3-5] showed that a large proportion of French people (46% to 71%) use the Internet to seek medical or health related information. Many people also use social media, such as forums, to communicate with others with the same health concerns and share information related to their illnesses, feelings, medication use and many other aspects [6], which offers promising opportunities for public health surveillance with a rich internet-based, patient-generated source.

The World Health Organization (WHO) defines Pharmacovigilance as "the science relating to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problems". It begins during clinical trials and continues after the drug is released onto the market.

However a study [7] showed that 60% of potentially fatal ADRs were not described in initial drug labels and 39% were not included in any report of randomized controlled trials. The main pharmacovigilance tools are spontaneous reporting systems, driven by drug agencies, like the U.S. FDA' (Food and Drug Administrations) Adverse Event Reporting System (FAERS) which gathers voluntary reports by healthcare professionals and consumers (59% by professionals Vs. 41% by consumers in 2006, and 46% Vs. 54% in Q1 2015 [8]). It can also include Phase IV clinical trials driven by pharmaceutical companies and governmental agencies [9]. Despite such systems, the underreporting of ADRs by the patients as well as by the health professionals remains a significant limitation [10;11].

Several studies have already demonstrated the value of mining ADR from social media posts [12-14]. However, in contrast to the numerous studies in social media, the potential of utilizing this data for pharmacovigilance has not yet been fully exploited. It represents only 0.5% of publications with "social media" (SM) keyword query in the PubMed database (Figure 1A). Figure 1B shows that the number of publications with "SM + pharmacovigilance" as keywords has increased exponentially in the last five years.

A recent scoping review [11] outlined five complete steps that should be taken for processing ADR extraction in social media: (1) data collection, (2) preprocessing, (3) entity recognition (for drugs and symptoms), (4) identifying the relationship between drug and symptom, and (5) results evaluation. Since content and language of medical social media differ from those of general social media and of clinical documents, specific text mining methods or techniques based on Natural Language Processing (NLP) are necessary for step (3) and step (4) in order to identify medical concepts (such as drugs, symptoms, etc.) and relations among them [15]. It is evident that the performance of the text mining methods plays a decisive role in ADR signal detection.

From a text mining perspective, the key challenge is that Internet users' expressions are usually informal and colloquial, especially when they describe their feelings and symptoms. However, researches have progressed using (i) various data sources, such as forum messages [16;17], Twitter micro blogs

[18;19] and Yahoo Wellness Groups [12], in (ii) different languages, (iii) diverse approaches, such as Support Vector Machine (SVM) and Conditional Random Fields (CRF) [20], and (iv) different medical terminologies, making it difficult to compare their performances [9].

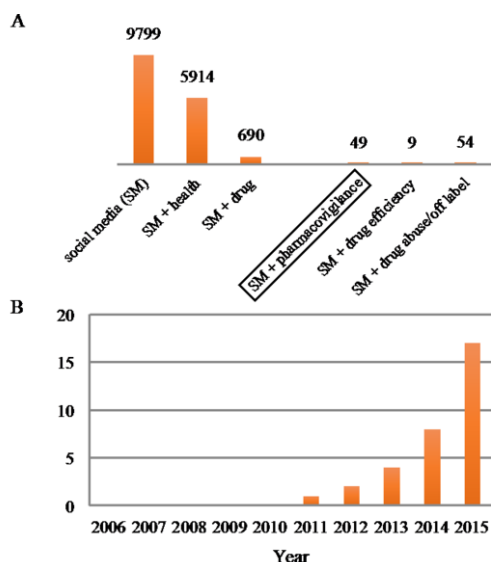


Figure 1 – (A) Number of publications with related key words (B) Distribution of “social media (SM) + pharmacovigilance” publications over 2006-2015

In this article, we introduce the approach developed in ADR-PRISM project in order to mine pharmacovigilance signals from social media. The utilized corpus, lexicons and methods of automated annotation are described in the next section. Moreover we highlight the evaluation protocol, which includes a detailed sample size determination and corpus constitution. The preliminary results are then discussed.

Materials and methods

Corpus

With the objective of extracting ADRs reported by patients on social media, we selected four French language, health related web sites:

- www.atoute.org
- forum.doctissimo.fr
- santemedecine.journaldesfemmes.com/forum/
- www.e-sante.fr/forums

These sites were selected through search engines and the CISMef web site, which is a catalog and index of French health resources on the Internet, and were evaluated using the Net scoring Tool [21].

An extractor targeting patients’ messages using the HTML structure was then applied to extract messages from these forums. The extractor uses the names of drugs as key words to identify all discussions quoting one of the key words, then extracts and cleans the messages from the discussions (removing useless information, like ads, signatures and quotations). We selected 50 drugs of interest as input key words, and for each of these drugs, 20 discussions were randomly picked and extracted. This extraction conducted to

the identification of 325 435 messages published between 2002 and 2014, corresponding to 967 distinct discussions.

The analysis showed that nearly 50% of the messages explicitly mentioned at least one symptom, and this ratio is in-between that of general forum posts (24%) [22] and that of drug reviews (80%) [23], which confirmed that our selection and evaluation of web sites and extraction were effective.

Lexicons

The thesaurus RacinePharma, which includes 5164 drug names, was used to identify drug mentions. This terminology is created by CISMef and *Service d’Informatique Biomédicale du CHU de Rouen* (SIBM) and updated monthly according to the French *Base de Données Publique des Médicaments* (BDPM), which ensures that it covers all medications on the French market that might be mentioned on social media.

Medical Dictionary for Regulatory Activities (MedDRA) version 15.1 was used to identify medical terms including symptoms, signs, diseases, diagnoses, names and results of analysis etc. In this article, we will use “symptom” to refer to all these terms in order to facilitate writing. MedDRA has a five level structure with a classification in 26 medical disciplines (SOC – System Organ Class), then in HLGT (High Level Group of Terms), HLT (High Level Terms), PT (Preferred Terms), and finally LLT (Lowest Level Terms). The coding of ADRs is done by the LLT, which includes synonyms, lexical variants, sub-elements, familiar expression or “old” terms, thus suits for our study. On this basis, we put in place a strategy to automatically overcome orthographic variations or missed/added terms into the LLT, thus built an extended version of MedDRA in order to increase the coverage without predicting all lay vocabulary in social media. When assessing the performance of automatic recognition of symptoms, we decided to consider the PT level, which clusters the synonyms or lexical variants that might be used by different posters. For example, if we select the PT “anxiety” to evaluate, recognitions of all LLT under this PT, like “worry”, “anxiety”, “anguish”, would be grouped and examined together. As MedDRA is a directed acyclic graph, there may exist multiple paths for the same entity. For example, the PT “scar” belongs to SOC “skin and subcutaneous tissue disorders” and also SOC “injury, poisoning and procedural complications”. In such situations, all possible hierarchies would be considered in the same manner.

Annotation

The Smart Taxonomy Facilitator (STF) Skill Cartridge™ developed by Expert System was applied on the initial corpus. It combines a rule-based approach and a dictionary-based approach. The latter includes two main technologies: (i), Fuzzy Term Matching, to take into account possible variants of the terms present in the taxonomy, thus reducing the number of false negatives; (ii) Relevance Scoring, which applies a series of heuristics that assigns a score to each extracted concept, and thus eliminates the least relevant concepts in order to reduce false positives. STF also exploits lexical labels (part-of-speech tagging) to address ambiguity issues.

We integrated in the Skill Cartridge the domain specific dictionaries (RacinePharma and MedDRA) and some internal rules established by our pharmacovigilance experts and text mining experts. Then the fuzzy matching parameters were adapted respectively for drug and symptom recognition.

ADR corresponds to a ternary relationship between (i) a patient and (ii) a symptom related with (iii) a drug through a causal relationship. We identified the linguistic patterns that corresponded to the five major semantic relations between

these three entities: administration (take, test, try, treatment, intake of, etc.), causal relationship (cause, give, result of, since, because of, etc.), sensation (suffer, feel, etc.), drug stop (stop to avoid, to arrest, etc.) and intolerance (endure, allergy, etc.). With the pre-defined linguistic patterns, the ADR Skill Cartridge™ is able to identify multiple relationships between one or more drugs and/or symptoms within one sentence.

Evaluation

Protocol overview

ADR mining from social media may have two different utilisations: (i) routine signal detection for public health and surveillance; and (ii) focused drug- (or symptom-) signal detection, mainly for pharmaceutical industry. We established a protocol to evaluate the performance of the recognition of drugs, symptoms and their relationships in those two contexts. We will describe in the next sections (1) the constitution of sub-corpus, i.e. selection of drugs and symptoms of interest and determination of sample size, (2) the establishment of gold standard, i.e. manual annotation and the guideline of manual annotation, (3) the statistics analysis for comparison.

Data sets

We expected to evaluate the general performance on all identified drug and symptom mentions, and also the performance on certain specific concepts. We therefore selected 12 drugs, including the most frequent in the corpus, the most sold in France in 2013, the most interesting according to the pharmacovigilance experts and we selected some drugs randomly. We selected 9 symptoms, based on similar principles, i.e., the most frequent, the most interesting and some randomly picked ones. For each selected concept, the sample size is calculated under the hypothesis of precision (or recall) = 0.5 ± 0.15 , with a significant level of 0.05. The posts containing at least one of the selected concepts were then pooled to build the sub-corpus for evaluation.

Manual annotation

An annotation guideline was established for human experts to annotate all words that refer to a drug or a symptom, and then annotate for every pair of drug and symptom whether there was a causal relationship between them in the context, without using any expert knowledge, experience or intuition to prejudice. Two pharmacovigilance experts with experience in ADR reporting, annotated blindly and independently a part of

the sub-corpus. Both experts annotated a common part of the messages, which aims to estimate the inter-annotator agreement, to perfect the guidelines and to improve the quality of manual annotation. We then considered this manual annotation as Gold Standard.

Comparison

The basic metrics used to evaluate the performance are precision, recall and their harmonic mean (F-measure). Three different result types are examined: false negative (FN) for non recognition of relevant terms, false positives (FP) for irrelevant positive recognitions and true positive (TP) for correct positive recognitions. The precision, recall and F-measure are defined respectively as Eq.1,

$$p = \frac{TP}{TP+FP}, \quad r = \frac{TP}{TP+FN}, \quad F_1 = \frac{2pr}{p+r} \quad (1)$$

These statistics were computed globally and also specifically for each selected concept.

Format and implementation

The extracted forum messages were transformed in XML format for applying the STF Skill Cartridge. Automated annotations in XML format were parsed with R 3.3.1 *xml2* packages. The sampling for sub-corpus constitution was carried out by building the list of message IDs with R and then integrating the corpus in the Skill Cartridge.

Results

Description of dataset

With the corpus described above,

- 55 777 entities of drug names from 34 265 messages have been annotated by the Skill Cartridge with thesaurus RacinePharma, which concern 1383 distinct drugs;
- 429 424 entities of symptoms from 153 995 messages have been annotated by the Skill Cartridge with thesaurus MedDRA, which concern 4861 distinct MedDRA terms.
- On the basis of drug and symptom recognitions, 1385 ADRs have been identified from 1129 messages.

Table 1 shows an overview of our dataset.

Table 1 – Dataset overview

Corpus	Mentions			Discussion	Messages	Messages containing			
	Drugs	Symptoms	ADRs			Drugs	Symptoms	Both	ADRs
Total	55777	424924	1385	967	325435 % of line total	34265 10,5%	153995 47,3%	27394 8,4%	1129 0,3%
Atoute	5457	36314	139	261	27234	2825 10,4%	11421 41,9%	2229 8,2%	109 0,4%
Doctissimo	15980	90876	432	565	73266	9134 12,5%	32577 44,5%	7122 9,7%	358 0,5%
E-sante	34221	297010	806	108	224946	22232 10,5%	109792 48,8%	17978 8,0%	656 0,3%
Sante-medecine	119	724	8	33	286	74 10,5%	205 71,7%	65 22,7%	6 2,1%

The 12 drugs used for evaluation included four categories:

1. from the most frequent drugs in the corpus, PUREGON (follitropin beta), SPASFON (phloroglucinol, trimethylphloroglucinol), and TARCEVA (erlotinib);

2. from the top 15 most sold in France, ASPIRINE (acetylsalicylic acid), LEVOTHYROX (levothyroxin), and DOLIPRANE (paracetamol/acetaminophen);
3. from the most interesting according to the pharmacovigilance experts: METHADONE, DIANE 35 (ethinylestradiol, cyproterone acetate) and PROZAC (fluoxetine);

4. three drugs chosen at random: IXEL (milnacipran), CHAMPIX (varenicline) and GLIVEC (imatinib).

For symptoms, the 3 most frequent PT selected were “anxiety”, “pain” and “fatigue”; the 3 of interest PT were “death”, “hypersensitivity” and “injury”; the 3 randomly selected PT were “basedow’s disease”, “moderate mental retardation” and “fungal infection”.

Gold standard

For drug and symptom recognition, the sample size for gold standard was determined with the method detailed in the previous section, which corresponds to 45 messages per concept. As one message can contain several occurrences of the same concept, and also mentions of other concepts, the sub-corpus for evaluation carries more entities than 45 per concept. Then for ADR detection, we evaluated all 1129 messages containing at least one ADR identified by the automatic approach. Two human experts read all messages in the sub-corpus and annotated the drugs, symptoms, and the relationships between them as being an ADR or not. The gold standard corpus, corresponding to 785 mentions of drugs and 908 mentions of symptoms, was then used for comparison.

Comparison results

The sub-corpus for evaluating drug name recognition corresponds to 561 messages, in which the Skill Cartridge identified 721 occurrences of drugs corresponding to 27 distinct drugs. Table 2 shows the global scores and the scores obtained for each drug in terms of precision, recall, and F-measure.

Table 2 – Evaluation results of drug name recognition

Drug	Precision	Recall	F-measure
USE CASE			
PUREGON	1,00	1,00	1,00
SPASFON	1,00	0,98	0,99
TARCEVA	1,00	1,00	1,00
ASPIRINE	1,00	1,00	1,00
LEVOTHYROX	0,92	1,00	0,96
DOLIPRANE	1,00	1,00	1,00
METHADONE AP-HP	0,95	1,00	0,98
DIANE	1,00	0,92	0,96
PROZAC	1,00	0,99	0,99
IXEL	1,00	1,00	1,00
CHAMPIX	1,00	0,92	0,96
GLIVEC	1,00	1,00	1,00
...
OVERALL	0,98	0,90	0,94

The sub-corpus for evaluating symptom recognition corresponds to 401 messages, in which the Skill Cartridge identified 640 mentions concerning 59 distinct PTs. The results are shown in Table 3. Although, we present in this table only the PT level the MedDRA hierarchy makes it possible to display similar results, at other levels (SOC, HLT, HLT).

Table 3 – Evaluation results of symptom recognition

MedDRA term (PT)	Precision	Recall	F-measure
USE CASE			
anxiety	0,98	0,86	0,92
pain	0,97	0,89	0,93
fatigue	0,98	0,97	0,98
death	0,99	0,97	0,98
hypersensitivity	1,00	0,92	0,96
Injury	0,92	0,96	0,94

basedow's disease	0,75	1,00	0,86
moderate mental retardation	1,00	1,00	1,00
Fungal infection	1,00	0,91	0,95
...
OVERALL	0,98	0,69	0,81

For ADR identification, we evaluated all messages containing ADR annotation and obtained a precision 0.78, recall 0.63 and F-measure 0.70. Taking into account the great challenges of the processing of social media texts, this relation detection result is encouraging and promising for downstream analysis.

Discussion

In this paper, we have described a methodology by which text messages on social media can be effectively transformed into a usable format for pharmacovigilance. The protocol of evaluation, which includes a detailed sample size determination, has been highlighted, which is often obscure in previously published works. We have obtained a nearly perfect accuracy on recognition of drug names, and good performance on recognition of symptoms and ADR relations. It seems that some types of mentions or relation patterns are easier to extract than others. One of the key issues is still the informal narrative in social media containing many grammatical errors, abbreviations, spelling mistakes and lay terminology.

Our performance of recognition (F-measure 0.94 for drugs and 0.81 for symptoms) is comparable with other studies in the domain. In CHEMDNER BioCreative IV challenge [24], the chemical compound and drug name recognition task reached an F-measure of 0.88, while the disease named entity recognition reached an F-measure of 0.86 in BioCreative V challenge [25; 26]. With Electronic Health Records (EHRs) data, the F-measure of drug name recognition varies from 0.73 to 0.89 [27]. With social media, although various studies have attempted to adapt different methods to this specific text data source, there is still a gap on recognition performance due to informal and colloquial expressions. Most pilot studies of mining ADRs from social media [28-30] have investigated for English language, and the F-measure ranged from 0.58 to 0.82. The performance depends mainly on the size and quality of dataset. A study of detecting drug effects from a Spanish health forum has obtained a precision of 0.48 and recall of 0.59 [31]. In French language social media, a study of automatic identification of drug-related medical conditions on drug review [23] obtained a F-measure of 0.95 for chemicals, 0.86 for signs/symptoms and 0.82 for diseases, however the relations are not considered in this work. Moreover their corpus and evaluation set are much smaller than ours.

Even if the automated annotation of ADR relations is now restricted to the co-occurrence of drugs and symptoms in the same sentence, our human annotators were asked to annotate all ADRs in the post regardless of the sentence boundary, which allows us to further assess the impact of the sentence restriction and eventually improve the performance of detection of relations across sentence boundaries. The method of evaluation presented in this article contains potentially a bias of overestimation of recall. The false negatives are actually underestimated due to the fact that we worked with messages containing at least one annotation for one of the selected drugs, symptoms, or one annotation of ADR. Rational for choosing this approach is that half of the messages did not exhibit any entity of interest (neither drug nor symptom).

Next step will be applying signal detection methods within the pharmacovigilance database issued from French social media. A comparison of these potential ADR signals with those

detected from traditional reporting data will be performed. More work remains to examine how social media data can be incorporated into overall pharmacovigilance systems.

Conclusion

Our approach provides very encouraging preliminary results of recognition of drug names, symptoms, and ADRs in social media texts, which offer a promising basis for downstream analysis of routine or specific ADR signal detection.

Acknowledgements

This work was labeled by the competitiveness cluster Cap Digital and funded by the DGE and territorial collectivities (Ile de France and Haute Normandie) under the 16th FUI (Fonds Unique Interministériel) request for proposal through the ADR-PRISM project.

References

- [1] R. Harpaz et al., "Text mining for adverse drug events: the promise, challenges, and state of the art," *Drug Saf.*, vol. 37, no. 10, pp. 777–790, Oct. 2014.
- [2] "Digital in 2016," *We Are Social Singapore*, 2016.
- [3] H. Romeyer, "La santé en ligne," *Commun. Inf. Médias Théories Prat.*, no. Vol. 30/1, Feb. 2012.
- [4] Conseil National de l'Ordre des Médecins, "Vers une meilleure intégration d'Internet à la relation médecins-patients," 2010. [Online]. Available: <https://www.conseil-national.medecin.fr/article/vers-une-meilleure-integration-d%E2%80%99internet-la-relation-medecins-patients-982>. [Accessed: 19-Apr-2016].
- [5] V. Gombaut, "Deux ménages sur trois disposent d'internet chez eux," *Insee Première*, Mar-2011. [Online]. Available: http://www.insee.fr/fr/themes/document.asp?ref_id=ip1340. [Accessed: 19-Apr-2016].
- [6] S. Katsahian et al., "Evaluation of Internet Social Networks using Net scoring Tool: A Case Study in Adverse Drug Reaction Mining," *Stud. Health Technol. Inform.*, vol. 210, pp. 526–530, 2015.
- [7] B. Seruga, L. Sterling, L. Wang, and I. F. Tannock, "Reporting of Serious Adverse Drug Reactions of Targeted Anticancer Agents in Pivotal Phase III Clinical Trials," *J. Clin. Oncol.*, vol. 29, no. 2, pp. 174–185, Jan. 2011.
- [8] "FAERS Reporting by Healthcare Providers and Consumers by Year," *FDA Adverse Events Reporting System (FAERS)*, 24-Nov-2015. [Online]. Available: <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm070456.htm>. [Accessed: 16-Dec-2016].
- [9] A. Sarker et al., "Utilizing social media data for pharmacovigilance: A review," *J. Biomed. Inform.*, vol. 54, pp. 202–212, Apr. 2015.
- [10] P. Tubert, B. Bégaud, J. C. Péré, F. Haramburu, and J. Lellouch, "Power and weakness of spontaneous reporting: a probabilistic approach," *J. Clin. Epidemiol.*, vol. 45, no. 3, pp. 283–286, Mar. 1992.
- [11] J. Lardon et al., "Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review," *J. Med. Internet Res.*, vol. 17, no. 7, Jul. 2015.
- [12] B. W. Chee, R. Berlin, and B. Schatz, "Predicting Adverse Drug Events from Personal Health Messages," *AMIA Annu. Symp. Proc.*, vol. 2011, pp. 217–226, 2011.
- [13] A. Yates and N. Goharian, "ADRTTrace: Detecting Expected and Unexpected Adverse Drug Reactions from User Reviews on Social Media Sites," in *Advances in Information Retrieval*, P. Serdyukov, P. Braslavski, S. O. Kuznetsov, J. Kamps, S. Rüger, E. Agichtein, I. Segalovich, and E. Yilmaz, Eds. Springer Berlin Heidelberg, 2013, pp. 816–819.
- [14] M. Abou Taam et al., "Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France," *J. Clin. Pharm. Ther.*, vol. 39, no. 1, pp. 53–55, Feb. 2014.
- [15] K. Denecke, "Extracting medical Concepts from medical social media with clinical NLP tools: a qualitative study," *Proc. Fourth Workshop Build. Evaluation Resour. Health Biomed. Text Process.*, 2014.
- [16] A. Yates, N. Goharian, and O. Frieder, "Extracting Adverse Drug Reactions from Forum Posts and Linking them to Drugs," *Proceeding 2013 ACM SIGIR Workshop Health Search Discov.*, Aug. 2013.
- [17] A. Benton et al., "Identifying potential adverse effects using the web: a new approach to medical hypothesis generation," *J. Biomed. Inform.*, vol. 44, no. 6, pp. 989–996, Dec. 2011.
- [18] J. Bian, U. Topaloglu, and F. Yu, "Towards Large-scale Twitter Mining for Drug-related Adverse Events," in *Proceedings of the 2012 International Workshop on Smart Health and Wellbeing*, New York, NY, USA, 2012, pp. 25–32.
- [19] K. Jiang and Y. Zheng, "Mining Twitter Data for Potential Drug Effects," in *Advanced Data Mining and Applications*, H. Motoda, Z. Wu, L. Cao, O. Zaiane, M. Yao, and W. Wang, Eds. Springer Berlin Heidelberg, 2013, pp. 434–443.
- [20] J. Lafferty, A. McCallum, and F. Pereira, *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. 2001.
- [21] "NetScoring : critères de qualité de l'information de santé sur l'Internet." [Online]. Available: <http://www.chu-rouen.fr/netscoring/>. [Accessed: 22-Apr-2016].
- [22] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on twitter? Mining tweets for adverse drug reactions," *AMIA Annu. Symp. Proc.*, vol. 2014, pp. 924–933, 2014.
- [23] M.-H. François, G. Cyril, and Z. Pierre, "Identification of Drug-Related Medical Conditions in Social Media," 2016.
- [24] M. Krallinger, F. Leitner, O. Rabal, M. Vazquez, J. Oyarzabal, and A. Valencia, "CHEMDNER: The drugs and chemical names extraction challenge," *J. Cheminformatics*, vol. 7, no. Suppl 1, p. S1, Jan. 2015.
- [25] C.-H. Wei et al., "Assessing the state of the art in biomedical relation extraction: overview of the BioCreative V chemical-disease relation (CDR) task," *Database J. Biol. Databases Curation*, vol. 2016, Mar. 2016.
- [26] Q. Wei, T. Chen, R. Xu, Y. He, and L. Gui, "Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks," *Database J. Biol. Databases Curation*, vol. 2016, 2016.
- [27] S. Liu, B. Tang, Q. Chen, and X. Wang, "Drug Name Recognition: Approaches and Resources," *Information*, vol. 6, no. 4, pp. 790–810, Nov. 2015.
- [28] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, and G. Gonzalez, "Towards Internet-age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts to Health-related Social Networks," in *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, Stroudsburg, PA, USA, 2010, pp. 117–125.
- [29] H. Sampathkumar, X. Chen, and B. Luo, "Mining Adverse Drug Reactions from online healthcare forums using Hidden Markov Model," *BMC Med. Inform. Decis. Mak.*, vol. 14, p. 91, 2014.
- [30] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn, and G. Gonzalez, "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features," *J. Am. Med. Inform. Assoc. JAMIA*, vol. 22, no. 3, pp. 671–681, May 2015.
- [31] I. Segura-Bedmar, P. Martínez, R. Revert, and J. Moreno-Schneider, "Exploring Spanish health social media for detecting drug effects," *BMC Med. Inform. Decis. Mak.*, vol. 15, no. 2, pp. 1–9, 2015.

Address for correspondence

Xiaoyi Chen
Inserm UMRS 1138 - Equipe 22
15 Rue de l'école de médecine 75006 Paris, France
Email: xiaoyi.chen@inserm.fr

Predicting Adverse Outcomes in Heart Failure Patients Using Different Frailty Status Measures

Yan Cheng^a, Yijun Shao^a, Charlene R. Weir^{b,c}, Rashmee U. Shah^d,
Bruce E. Bray^b, Jennifer H. Garvin^b, Qing Zeng-Treitler^{a,e}

^a Biomedical Informatics Center, George Washington University, Washington, DC, USA,

^b Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA,

^c Veterans Affairs Medical Center, Salt Lake City, UT, USA,

^d Division of Cardiovascular Medicine, University of Utah, Salt Lake City, UT, USA,

^e Washington DC VA Medical Center, Washington, DC, USA

Abstract

Frailty is an important outcome predictor in older patients. We randomly sampled 12,000 veterans with heart failure diagnosed in 2010. The topic modeling method was applied to identify frailty-related topics from the clinical notes in the electronic medical records. The frailty topics were classified into five deficit areas including physical functioning (PF), role-physical (RP), general health (GH), social functioning (SF), and mental health (MH). We experimented with different covariates and four different frailty measures: individual frailty topics, number of distinct frailty topics, a dichotomous deficit category, and the number of distinct deficits, respectively. A total of 8,531 (71.1%) patients had at least one frailty topic. The prevalence of GH, PF, MH, SF, and RP deficits were 89.0%, 61.3%, 56.9%, 40.6%, and 9.5%, respectively. PF deficits (yes/no) and the number of distinct deficits were the most consistent, significant predictors of adverse outcomes of rehospitalization or death.

Keywords:

Medical Informatics; Frail Elderly

Introduction

Frailty commonly occurs in older adults and is an important determinant of health outcomes [3; 14; 16; 26]. Frailty is distinct from comorbidity and disease, and is a multifaceted combination of fatigue, weakness, malnutrition, and mobility [5, 15]. Frailty not only affects patients and caregivers, but is also a leading indicator of worsening health outcomes, including death. It is a barometer of how well patients may respond to treatment. However, frailty measurements are rarely collected in a quantitative, reliable fashion in routine patient care.

Frailty is specifically an important measure to assess cardiac peri-operative risk, morbidity, and mortality. Older patients with heart failure (HF) comprise a growing proportion of the cardiac surgery population. The Society for Thoracic (STS) risk model is often used to estimate risk, but includes no variables directly related to patient frailty [1]. Recently, smaller studies have demonstrated the added predictive value of frailty measures, for example the 5-minute gait speed [2] and a comprehensive frailty assessment [20]. Still, frailty is a key dimension that continues to be absent from quantitative risk prediction, in part because it is challenging to capture this metric on a large scale.

Fortunately, clinicians commonly document various aspects of frailty in clinical notes, especially when treatment plans change and/or when a patient's quality of life is impacted, suggesting that perceptions of frailty is a component of clinician's mental model of the patient's status. The Department of Veteran Affairs (VA) national electronic health record (EHR) database is a particularly rich data source with an extremely large, older patient population with a comprehensive collection of different types of clinical notes.

In this study, we expanded on prior work examining the association between the number of frailty topics and adverse outcomes. We used existing frailty assessment instruments for ontological guidance, and created four different measures of frailty status to evaluate their role in predictive modeling.

Prior Work

In our prior work [19], we used a case control study design to estimate the association of the number of frailty with poor outcomes among HF patients. The outcome of interest was ≥ 2 hospitalizations following index HF hospitalization or death. A total of 709,389 notes were included from 12,000 patients. For training, we randomly selected 50,000 notes from 4,000 patients with the outcome, and 50,000 notes from 8,000 patients without the outcome. We ran the latent Dirichlet allocation (LDA) program from a java software package called MACHINE Learning and Language Toolkit (MALLET), on the 100,000 notes, with the initial number of topics set to be 700.

To identify stable topics that were consistently present in different LDA runs, we first independently applied each of the 3 learned LDA models to the 709,389 notes using the topic inference tool included in MALLET. This step yielded 3 topic proportions per note for each note. Next, we defined a stable topic to be present in a note if at least 2 of the 3 topics in the topic triple corresponding to the stable topic had a proportion of 0.02 or higher in that note. This step produced 556 stable topics. An informatician and a physician independently reviewed all the stable topics and identified 53 topics that were related to frailty. The inter-rater agreement between reviewers measured by Kappa was found to be 0.818. We used the labels assigned by the informatics expert as human interpretations.

We discovered that increased number of frailty topics was statistically associated with increased risk of poor outcomes. Each additional frailty topic was associated with a 7% higher

risk of adverse outcomes. Compared to patients with <4 frailty topics, those with ≥ 4 frailty topics had two times greater risk of developing an adverse outcome within 1 year following the initial HF diagnosis.

While the results were promising, further studies are needed because the area under ROC curve (AUC) was suboptimal in the prior work, at just above 0.6 regardless of the way we parameterized the frailty variable (coding it as a continuous or binary variable). This is partly because only a small number of covariates were used. However, when we included a larger number of covariates, the statistical significance of the number of frailty topics was diminished. Thus, more robust frailty measurements are needed to predict adverse outcomes.

Methods

Data Source

In this study, we used the Veterans Administration Informatics and Computing Infrastructure (VINCI) as the data source. VINCI contains comprehensive patient health and medical information from the US nationwide veterans' EMR, which include both structured (i.e., race, gender, diagnosis code) and unstructured data (data in text documents, i.e., clinical notes).

Study Population

Patients in our study population were the same as those identified in the prior work [19]. They were 12,000 randomly sampled veterans with one International Classification of Disease 9th Clinical Modification (ICD-9-CM) HF diagnosis of 428.0-428.9 in 2010. They were composed of 4,000 veterans who experienced death or ≥ 2 HF-caused hospitalizations during the year after the first HF diagnosis and 8,000 veterans who did not experience death and had at most 1 HF hospitalization during the year after first HF diagnosis.

Outcomes and Predictors

The main outcome of this study was defined as ≥ 2 all-cause hospitalizations or death within 1 year after the first diagnosis of HF.

The predictors and covariates were identified from both structured and unstructured data. For the structured data, we used patient birthdate, gender, and ICD-9-CM diagnoses at each visit. Age was calculated as baseline age at the first diagnosis of HF. Patients' comorbidities represented by ICD codes during the year before the first HF diagnosis were captured. The Charlson Comorbidity Index (CCI) was calculated based on these ICD codes via the methods described by Quan *et al.* [18].

For the unstructured data, we extracted topics from the Text Information Utility (TIU) notes. All the TIU notes dated within one year prior to the first HF diagnosis were extracted. Frailty indicators, which were not available in the structured data, were extracted from the TIU notes using the topic modeling technique.

Frailty Measurement Development

As described in the introduction section, we identified 53 frailty topics. Using these topics, we created four types of frailty measurements using the ontology knowledge from SF-36 and Frailty Index [23]. SF-36 and Frailty Index are commonly used among a number of frailty assessment instruments. We grouped the frailty topics into five deficit domains (called "deficits"), including physical functioning (PF), role-physical (RP), general health (GH), social functioning (SF), and mental health (MH). Topics relating to

deficits in physical activities were grouped as PF; topics relating to deficits in role activities were grouped as RP; topics relating to general health perception or vitality (energy and fatigue) were grouped as GH; topics relating to deficits in social activity were grouped as SF; and topics relating to mental health were grouped as MH. The purpose of grouping is to see if some deficit domains are better than others to predict adverse outcome and if grouped deficits are more predictive than individual frailty topics. These comparisons are critical for developing frailty ontology in our future study. Since we are investigating how to measure frailty, each frailty topic and frailty deficit domain were first treated as individual variables. We then calculated an aggregate of the total number of distinct frailty topics and the total number of distinct deficits as additional variables.

Logistic Regression Models

We used logistic regression models to analyze the association of frailty and the outcome at the individual level. We assessed each of the four frailty measures (individual frailty topics, number of distinct frailty topics, deficit category, and number of distinct deficits, respectively) as predictors for the outcome. We also experimented with 4 different sets of covariates: set #1 including age, gender, and CCI; set #2 including age, gender, and individual comorbid conditions identified by ICD-9 diagnoses; set #3 including age, gender, individual comorbid conditions, and individual topics not related to frailty; and set #4 including all structured data (age, gender, CCI, ICD-9 diagnosis, Current Procedure Terminology [CPT] procedure codes, medications, and medical note type) and non-frailty topics. For each set of covariates, we built four logistic regression models separately with including one of four frailty measures in each model. Given the large number of comorbid conditions and topics, we used automatic stepwise selection methods to set entry p-value at 0.2 and p-value at 0.05 to choose predictors other than the four frailty measures. A multi-linearity test was also conducted for each model to make sure there was no variable with variance inflation factor (VIF) higher than 10.

Results

Demographics

Among the studied cohort, 97.9% were males. The mean age was 69.7 (SD 12.1) years old, with 48.9% veterans aged 60-69, 26.1% aged 70-79, 21.9% aged 80-89, and 3.1% aged 90 or older. Just under a third of veterans (31.4%) had CCI of 0 (based on ICDs dated within one year of the first HF diagnosis) indicating no comorbid conditions, 34.3% veterans had CCI of 1-2, and 34.3% veterans had CCI of 3 or above.

Frailty Measurements and Prediction Models

Among the total 53 frailty topics, 22 were grouped as PF deficits, 19 as GH deficits, 7 as MH deficits, 4 as SF deficits, and one as a RF deficit. The grouping of frailty topics is described in Table 1. A total of 8,531 (71.1%) patients had at least one frailty topic in their medical notes, among which 89.0%, 61.3%, 56.9%, 40.6%, and 9.5% had frailty topics in GH, PF, MH, SF, and RP, respectively.

As shown in Table 2 a-d, we created four sets of predictive models with each set using the same covariates and one of the four frailty measures as predictors, respectively. For each set of models, when the same covariates were used, the accuracy was very similar across the models, with AUC of 0.66, 0.80, 0.81, and 0.86 for set #1, #2, #3 and #4 of models, respectively.

Frailty Topics

Using individual frailty topics (rather than grouped into larger domains) sometimes resulted in a small (0.002 in AUC) improvement in prediction accuracy when compared with the other three frailty measures. As we included more covariates from set #1 to #4, the AUC improved from 0.66 to 0.87, regardless of what frailty measure we used. Including more covariates also caused fewer individual frailty topics to be significant (decreased from 12 in table 2 a to 2 in table 2 d). Most frailty topics did not consistently predict the outcome.

The topic variable “503;383;345” was the only one that was always significantly associated with increased risk of outcome. This variable was a topic related to physical functioning with top 5 key words of “point patient assistance bathing independence.” This is because self-care activity is an important risk factor for hospitalization or death, which is not captured in ICDs.

Deficits

The number of distinct deficits was consistently associated with the outcome and the association was always significant except for the model including age, gender, and ICD covariates. Compared to the number of distinct frailty topics, the number of distinct deficits was more predictive and significant. The odds ratio estimate of 1.02 to 1.32 for the number of distinct deficits were consistently higher than the odds ratio estimate of 0.99 to 1.08 for the number of distinct frailty topics in each corresponding set of models.

Among individual deficit variables, the PF was the only deficit that was consistently associated with increased risk of outcome. The GH was not predictive in any model; other deficits were significantly associated with the outcome occasionally depending on what covariates were included in the model..

When all structured data and other non-frailty topics were controlled, one additional deficit would increase risk of adverse outcomes by 6%. Although 6% seems small, the effect magnitude may be larger since the number of distinct deficits can go from 0 to 5. For example, compared to patients with 0 deficit, the risk of adverse outcomes would be increased by 12% and 34% among patients with 2 and 5 deficits, respectively.

Discussion

In this study, 71.1% veterans had at least one frailty topic in their medical notes. Since our cohort consisted of patients with HF, the prevalence of frailty was expected to be relatively higher than the general population. It also indicates the centrality of frailty in the mental models of cardiologists. Our findings are congruent with other work in frailty. A review study found that the prevalence of frailty ranged from 15-74%, depending on the study population and method of assessment [22].

This study shows the importance of using ontological knowledge to aggregate frailty findings. Individual frailty topics and the number of distinct topics are not the best predictors, when compared to the burden from aggregated deficits. Frailty is an aggregation of deficits as others have noted. When using frailty topics as individual variables, most topics became insignificant, because of the lower prevalence and because frailty cannot be determined by a single finding. When aggregated, the impact of frailty became more pronounced. For example, patients with ≥ 3 deficits had 1.16 times risk to have adverse outcomes compared with those with < 3 deficits.

Frailty is not the only predictor of outcome. Depending on the covariates included in the models, the AUC ranged from 0.659-0.667 to 0.868-0.870. More covariates resulted in higher AUC, and once individual ICDs instead of CCI were included in the model, then AUC improved substantially. According to the generally accepted rule for AUC, models adjusting for set #1 covariates were inferior to good, since the AUC was less than 0.7 but higher than 0.5; models adjusting for set #2, #3, or #4 covariates were strong, since the AUC were around 0.8 or above [10].

Many studies have reported that frailty is independently associated with a higher risk of death and other adverse outcomes among people with cardiovascular disease or the elderly in both short term and long term [6; 8; 11; 13; 21]. These studies in general reported frailty prediction models with lower accuracy than ours, regardless of frailty measure methods. These studies also used many different instruments to measure frailty including frailty index, modified frailty index, frailty-related phenotype, frailty criteria, frailty survey, and frailty related signs and symptom scales, to predict adverse outcomes [4;7;9;24;25]. The AUC of these prediction models were in the range of 0.55-0.77, which was lower than those of our models [4; 7; 9; 24; 25].

Among all the deficits, physical functioning was the only persistent predictor of the outcomes. Frailty, however, goes beyond physical functioning. The number of distinct deficits was robustly significant even when we included many other predictors besides age, gender, and ICDs. These findings suggest that the physical deficits are possibly related to a generalized functional status latent variable and perhaps an indicator of the patient’s overall burden. Our finding fills a research gap in understanding what deficit of frailty topics and if increased number of deficits would be significantly associated with the higher risk of the adverse outcomes, which remained unknown in the previous studies [12; 17].

Our study has some limitations. First, we used the most commonly used HF ICD-9-CM codes to identify the study population, which might miss patients that should be included otherwise. Second, this study explored a simple ontology of frailty by classifying the frailty topics according to patients’ physical, mental, social function and general health status. Third, we set an equal weight for different deficits, which may not be optimal, since these deficits may have variable effect magnitudes on adverse outcomes. Fourth, in this study, we focused on the overall outcome, not the outcome following specific treatments, which might be more useful.

Our final goal is to create a more detailed ontology to improve the identification of deficits and guide the aggregation of deficits into meaningful frailty levels. In the future studies, we will conduct a survey among HF patients to identify deficits from their perspective. We will evaluate patients’ frailty levels following major cardiac procedures and examine the association between frailty level and time-to-events.

Conclusions

In summary, aggregate frailty deficit measurements created based on frailty topics and ontology knowledge significantly and strongly predicted adverse outcomes among heart failure patients.

Table 1 Deficit and Key Words for Frailty Topics Variables

Deficit	Topics	Top Five Keywords
PF	133;132;102	fall risk score morse gait
	698;229;28	bed call reach light position
	24;144;324	walker cane gait wheelchair ambulation
	488;37;4	mobility functional goals balance transfers
	520;470;564	risk fall falls patient high
	547;20;641	resident fall bed lock risk
	419;394;578	ot adl dressing functional shower
	503;383;345	point patient assistance bathing independence
	439;104;250	assistance independent level mobility bowel
	437;218;399	feces shift output bm fall
	484;169;492	shoes wear issued fit size
	144;213;206	ability position eats discomfort occasionally
	543;231;663	staff minutes tolerated support set-up
	98;577;120	patient place fall assistance bed
	576;455;230	assist care cc independent bladder
	653;177;191	assessment risk fall patient sounds
	514;589;654	assist independent shift needed care
	366;588;404	provided support assist activity performance
	571;68;644	provided activity support adl occur
	529;422;226	risk patient fall assessment skin
RP	325;553;336	patient fall pain assess monitor
	472;70;490	fall risk patient assessment bed
GH	74;620;477	activities leisure activity group social
	194;401;354	bed resting monitor noted distress
	303;135;385	home care services health va
	540;605;131	skin dry intact bed warm
	206;499;637	appointment scheduled clinic show letter
	335;303;687	family patient member significant members
	247;171;303	call message left phone called
	249;359;139	care patient needed diet rehabilitation
	413;170;359	procedure sedation patient performed consent
	591;375;661	diet eat foods meals milk
	47;536;339	long term short rehab patient
	205;88;18	care patient days moving behaviors
	466;560;35	visit hbpc home caregiver medication
	157;120;289	good poor fair appetite time
	643;411;9	comments additional date staff assistance
	344;639;167	patient current condition complain detailed
	387;148;660	low diet high fat salt
	553;380;575	problems feeling trouble days difficult
	607;400;510	staff provided patient shift mod
	169;292;285	activity staff occur resident assistance
SF	693;666;317	daughter son home called spoke
	530;507;467	writer veteran stated contact contacted
	130;509;628	wife time son house years
	181;108;203	patient caregiver action spouse criteria
MH	225;99;107	alert oriented x3 distress acute
	90;361;691	oriented alert place time status
	36;78;476	memory cognitive average speech evaluation
	54;469;320	suicide risk thoughts plan suicidal
	106;221;92	suicidal ideation homicidal hallucinations mood
	281;604;604	restraint restraints patient family behavior
	158;567;319	wife home states dementia husband

Table 2 a – Prediction Performance of Four Different Measures of Frailty Topics With Set #1 Coveriates*

	Variable	OR	95% CI	AUC
Frailty Topics	133;132;102	1.37	1.21-1.56	0.667
	520;470;564	1.17	1.04-1.31	
	547;20;641	1.38	1.11-1.73	
	503;383;345	1.46	1.24-1.71	
	194;401;354	1.27	1.08-1.49	
	303;135;385	1.23	1.07-1.41	
	206;499;637	1.19	1.08-1.31	
	387;148;660	0.81	0.73-0.91	
	693;666;317	1.41	1.24-1.61	
	530;507;467	1.20	1.05-1.37	
Deficits	90;361;691	1.20	1.05-1.38	0.660
	106;221;92	1.23	1.07-1.42	
	#Distinct Frailty Topics	1.08	1.07-1.09	
	PF	1.61	1.47-1.76	
	RP	1.18	1.00-1.39	
	GH	0.95	0.86-1.05	
	SF	1.50	1.36-1.65	
	MH	1.38	1.26-1.52	

#Distinct Deficits	1.32	1.29-1.36	0.659
--------------------	------	-----------	-------

*Set #1 covariates: age, gender, and CCI

Table 2 b – Prediction Performance of Four Different Measures of Frailty Topics With Set #2 Covariates*

	Variable	OR	95% CI	AUC
Frailty Topics	419;394;578	0.78	0.63-0.97	0.801
	503;383;345	1.35	1.12-1.62	
	472;70;490	0.63	0.41-0.97	
	303;135;385	1.21	1.03-1.41	
	466;560;35	1.39	1.02-1.90	
	387;148;660	0.81	0.72-0.92	
Deficits	693;666;317	1.19	1.03-1.39	0.798
	#Distinct Frailty Topics	0.99	0.98-1.00	
	PF	1.15	1.04-1.28	
	RP	0.75	0.61-0.91	
	GH	0.91	0.82-1.01	
	SF	1.04	0.93-1.16	
#Distinct Deficits	MH	1.07	0.97-1.19	0.798
	#Distinct Deficits	1.02	0.99-1.05	

*Set #2 covariates: age, gender, and individual comorbid conditions

Table 2 c – Prediction Performance of Four Different Measures of Frailty Topics With Set #3 Covariates*

	Variable	OR	95% CI	AUC
Frailty Topic	503;383;345	1.41	1.17-1.71	0.816
	466;560;35	1.42	1.02-1.96	
Deficits	#Distinct Frailty Topics	1.01	0.99-1.02	0.815
	PF	1.25	1.12-1.40	
	RP	0.86	0.70-1.07	
	GH	1.03	0.91-1.15	
	SF	1.06	0.94-1.20	
	MH	1.13	1.01-1.27	
#Distinct Deficits	#Distinct Deficits	1.10	1.05-1.15	0.814

*Set #3 covariates: age, gender, individual comorbid conditions, and non-frailty topics

Table 2 d – Prediction Performance of Four Different Measures of Frailty Topics With Set #4 Covariates*

	Variable	OR	95% CI	AUC
Frailty Topic	503;383;345	1.45	1.18-1.78	0.870
	466;560;35	1.47	1.02-2.11	
Deficits	#Distinct Frailty Topics	1.00	0.98-1.01	0.869
	PF	1.21	1.07-1.38	
	RP	0.87	0.69-1.09	
	GH	1.00	0.88-1.14	
	SF	1.09	0.95-1.24	
	MH	1.02	0.90-1.16	
#Distinct Deficits	#Distinct Deficits	1.06	1.01-1.11	0.868

*Set #4 covariates: all structured data (age, gender, CCI, ICD-9 diagnosis, Current Procedure Terminology [CPT] procedure codes, medications, and medical note type) and non-frailty topics

Acknowledgements

This work is funded by the NIH grant R56 AG052536-01A1 and grants from the US Department of Veterans Affairs, Office of Research and Development, Health Services Research and Development including CHIR HIR 08-374, HIR 08-204, CRE 12-315 and the CREATE: A VHA NLP Software Ecosystem for Collaborative Development and Integration.

References

- [1] Short-Term Risk Calculator and Models, in: *Quality, Research & Patient Safety*, The Society of Thoracic Surgeons, 2016.
- [2] J. Afialo, M.J. Eisenberg, J.F. Morin, H. Bergman, J. Monette, N. Noiseux, L.P. Perrault, K.P. Alexander, Y. Langlois, N. Dendukuri, P. Chamoun, G. Kasparian, S. Robichaud, S.M. Gharacholou, and J.F. Boivin, Gait speed as an incremental predictor of mortality and major morbidity in elderly patients undergoing cardiac surgery, *J Am Coll Cardiol* **56** (2010), 1668-1676.
- [3] F. Buckinx, Y. Rolland, J.-Y. Reginster, C. Ricour, J. Petermans, and O. Bruyère, Burden of frailty in the elderly population: perspectives for a public health challenge, *Archives Of Public Health = Archives Belges De Santé Publique* **73** (2015), 19-19.
- [4] S. Buettner, D. Wagner, Y. Kim, G.A. Margonis, M.A. Makary, A. Wilson, K. Sasaki, N. Amini, F. Gani, and T.M. Pawlik, Inclusion of Sarcopenia Outperforms the Modified Frailty Index in Predicting 1-Year Mortality among 1,326 Patients Undergoing Gastrointestinal Surgery for a Malignant Indication, *J Am Coll Surg* **222** (2016), 397-407 e392.
- [5] A. Clegg, J. Young, S. Iliffe, M.O. Rikkert, and K. Rockwood, Frailty in elderly people, *Lancet* **381** (2013), 752-762.
- [6] J. Cockburn, M.S. Singh, N.H. Rafi, M. Dooley, N. Hutchinson, A. Hill, U. Trivedi, A. de Belder, and D. Hildick-Smith, Poor mobility predicts adverse outcome better than other frailty indices in patients undergoing transcatheter aortic valve implantation, *Catheter Cardiovasc Interv* **86** (2015), 1271-1277.
- [7] I. Drubbel, N.J. de Wit, N. Bleijenberg, R.J. Eijkemans, M.J. Schuurmans, and M.E. Numans, Prediction of adverse health outcomes in older people using a frailty index based on routine primary care data, *J Gerontol A Biol Sci Med Sci* **68** (2013), 301-308.
- [8] S.J. Evans, M. Sayers, A. Mitnitski, and K. Rockwood, The risk of adverse outcomes in hospitalized older patients in relation to a frailty index based on a comprehensive geriatric assessment, *Age Ageing* **43** (2014), 127-132.
- [9] D.B. Hogan, E.A. Freiheit, L.A. Strain, S.B. Patten, H.N. Schmaltz, D. Rolfson, and C.J. Maxwell, Comparing frailty measures in their ability to predict adverse outcome among older residents of assisted living, *BMC Geriatr* **12** (2012), 56.
- [10] D.W. Hosmer and S. Lemshow, *Applied Logistic Regression (2nd Edition)*, John Wiley & Sons, New York, NY, 2000.
- [11] J. Karam, A. Tsiouris, A. Shepard, V. Velanovich, and I. Rubinfeld, Simplified frailty index to predict adverse outcomes and mortality in vascular surgery patients, *Ann Vasc Surg* **27** (2013), 904-908.
- [12] D.H. Kim, C.A. Kim, S. Placide, L.A. Lipsitz, and E.R. Marcantonio, Preoperative Frailty Assessment and Outcomes at 6 Months or Later in Older Adults Undergoing Cardiac Surgical Procedures: A Systematic Review, *Ann Intern Med* **165** (2016), 650-660.
- [13] D.H. Kim and S. Schneeweiss, Measuring frailty using claims data for pharmacoepidemiologic studies of mortality in older adults: evidence and recommendations, *Pharmacoepidemiol Drug Saf* **23** (2014), 891-901.
- [14] L. Lahousse, B. Maes, G. Ziere, D.W. Loth, V.J.A. Verlinden, M.C. Zillikens, A.G. Uitterlinden, F. Rivadeneira, H. Tiemeier, O.H. Franco, M.A. Ikram, A. Hofman, G.G. Brusselle, and B.H. Stricker, Adverse outcomes of frailty in the elderly: the Rotterdam Study, *European Journal Of Epidemiology* **29** (2014), 419-427.
- [15] F. Lally and P. Crome, Understanding frailty, *Postgraduate Medical Journal* **83** (2007), 16-20.
- [16] J.E. Morley, B. Vellas, G.A. van Kan, S.D. Anker, J.M. Bauer, R. Bernabei, M. Cesari, W.C. Chumlea, W. Doehner, J. Evans, L.P. Fried, J.M. Guralnik, P.R. Katz, T.K. Malmstrom, R.J. McCarter, L.M. Gutierrez Robledo, K. Rockwood, S. von Haehling, M.F. Vandewoude, and J. Walston, Frailty consensus: a call to action, *Journal Of The American Medical Directors Association* **14** (2013), 392-397.
- [17] K. Oakland, R. Nadler, L. Cresswell, D. Jackson, and P.A. Coughlin, Systematic review and meta-analysis of the association between frailty and outcome in surgical patients, *Ann R Coll Surg Engl* **98** (2016), 80-85.
- [18] H. Quan, V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J.C. Luthi, L.D. Saunders, C.A. Beck, T.E. Feasby, and W.A. Ghali, Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data, *Med Care* **43** (2005), 1130-1139.
- [19] Y.M. Shao, A.F.; Ahmed, A.; Weir C.; Bray, B.E.; Shah R.; Redd, D.; Treitler Zeng Q., Identification and Use of Frailty Indicators from Text to Examine Associations with Clinical Outcomes Among Patients with Heart Failure, in: *AMIA 2016 Annual Symposium*, Chicago, IL 2016.
- [20] S. Sundermann, A. Dademasch, J. Praetorius, J. Kempfert, T. Dewey, V. Falk, F.W. Mohr, and T. Walther, Comprehensive assessment of frailty for elderly high-risk patients undergoing cardiac surgery, *Eur J Cardiothorac Surg* **39** (2011), 33-37.
- [21] S.H. Sundermann, A. Dademasch, B. Seifert, H. Rodriguez Cetina Bieffer, M.Y. Emmert, T. Walther, S. Jacobs, F.W. Mohr, V. Falk, and C.T. Starck, Frailty is a predictor of short- and mid-term mortality after elective cardiac surgery independently of age, *Interact Cardiovasc Thorac Surg* **18** (2014), 580-585.
- [22] I. Uchmanowicz, M. Lobo-Rudnicka, P. Szela, B. Jankowska-Polanska, and K. Lobo-Rudnicka, Frailty in heart failure, *Curr Heart Fail Rep* **11** (2014), 266-273.
- [23] J.E. Ware, Jr. and C.D. Sherbourne, The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection, *Med Care* **30** (1992), 473-483.
- [24] I.S. Widagdo, N. Pratt, M. Russell, and E.E. Roughead, Predictive performance of four frailty measures in an older Australian population, *Age Ageing* **44** (2015), 967-972.
- [25] J. Woo, J. Leung, and J.E. Morley, Comparison of frailty indicators based on clinical phenotype and the multiple deficit approach in predicting mortality and physical limitation, *J Am Geriatr Soc* **60** (2012), 1478-1486.
- [26] Q.-L. Xue, The frailty syndrome: definition and natural history, *Clinics In Geriatric Medicine* **27** (2011), 1-15.

Address for Correspondence

Qing Zeng-Treitler

zengq@email.gwu.edu

A Data-Driven Decision-Support Tool for Population Health Policies

Michal Chorev^a, Lavi Shpigelman^{a*}, Peter Bak^b, Avi Yaeli^b,
Edwin Michael^c, Ya'ara Goldschmidt^a

^a IBM Research - Haifa, Israel

^b IBM Watson Health, Haifa, Israel

^c University of Notre Dame, Indiana, USA

* Equal contribution first author

Abstract

Epidemiological models are key tools in assessing intervention policies for population health management. Statistical models, fitted with survey or health system data, can be combined with lab and field studies to provide reliable predictions of future population-level disease dynamics distributions and the effects of interventions. All too often, however, the end result of epidemiological modeling and cost-effectiveness studies is in the form of a report or journal paper. These are inherently limited in their coverage of locations, policy options, and derived outcome measures.

Here, we describe a tool to support population health policy planning. The tool allows users to explore simulations of various policies, to view and compare interventions spanning multiple variables, time points, and locations. The design's modular architecture, and data representation separate the modeling methods, the outcome measures calculations, and the visualizations, making each component easily replaceable. These advantages make it extremely versatile and suitable for multiple uses.

Keywords:

Computer-Assisted Decision Making; Statistical Models; Stochastic Processes

Introduction

An increasing number of epidemiological studies recognize that the front lines for tackling disease control and elimination are first and foremost policy and decision-making. Disease dynamics are heavily imprinted with uncertainty, complexity, and spatial heterogeneity [1-5]. This is particularly challenging for current strategies aimed at eliminating complex vector-borne diseases, such as malaria and lymphatic filariasis, which exhibit a high degree of geographic heterogeneity in infection patterns, and hence in transmission and extinction dynamics [6-9].

Indeed, previous works have shown that observed infection patterns display a high sensitivity to spatial and temporal factors. The dynamics of parasite transmission and extinction, as well as the intensity and duration of interventions, can vary significantly between communities owing to variations in initial ecological conditions [10-11].

Under these limitations, interventions must incorporate practices that effectively account for the variations in localized infection and control dynamics. This, in turn, underlines the need for new computational modeling tools. In contrast to

previous models, which were based on the assumption of spatial constancy in rate parameters [12], the new models must support the elimination of diseases within diverse endemic settings.

A critical requirement for these tools is the development of an effective interface to help decision-makers understand the processes driving disease transmission and the effects of interventions in different settings. Such a system should also be able to assess the robustness of the conclusions derived from model simulations. Moreover, it should have the ability to visualize different control scenarios to aid effective policy evaluation in a spatial context. The policy evaluation process should assist in the task of simultaneously optimizing several incommensurable and competing objectives. For example, these may include the desire to minimize the time it takes to eradicate a disease, the number of people affected by the disease and by the interventions, the costs for the health system, and the costs for the population. Any solution would compromise a tradeoff between objectives, but some solutions may be better with respect to others.

We present a tool that couples observed and modeled data within an interactive visualization environment. The tool enables policy makers to analyze and interpret outputs of simulations, map and identify areas with persistent infection, and run scenarios to evaluate the application of spatially variable intervention options to disrupt transmission reliably everywhere.

Our tool's aim is to allow policy planners to explore and compare simulated policies through a consumable interface that simplifies the presentation of spatio-temporal data under different intervention scenarios. The tool allows practical comparison of policies effect, and is easily customizable to specific regions or locations, time frames, and models.

Methods

Web Application Implementation

The tool is a web application that follows the standard Model View Controller (MVC) design pattern. We implemented the tool's front-end in HTML5 and JavaScript using jQuery (<https://jquery.com/>) and AngularJS (<https://angularjs.org/>). For visualization purposes, we used the following libraries: Bootstrap (<http://getbootstrap.com/>), D3 (<https://d3js.org/>), Leaflet (<http://leafletjs.com/>), and IBM's time-oriented visual analytics kit (see below), which was developed by our team. The server side was implemented in Python using the Flask web framework (<http://flask.pocoo.org/>). The database is

NoSQL sitting on a Couchbase server (<http://www.couchbase.com/>).

The visual analytics kit is a general-purpose visualization tool that is tailored to the analysis of time-series data by data scientists in healthcare. It is a JavaScript library that was developed separately from the tool described in this work [13] and is used here to display time series charts. Its primary goal is to shorten the time it takes to visualize and gain insights from data, without necessarily requiring knowledge in programming or in visualization per se. Users of the kit can create interactive charts and views of their data with a convenient declarative interface. In addition to the native JavaScript web-page integration capability, the library has interfaces for several different languages such as R, SAS, and Python, and can be used within Jupyter notebooks to display interactive figures. Among the available charts, users can find timelines, flow-charts, and matrix-views for many different tasks and purposes.

The focus of all of the kit's charts is the assessment of the temporal aspects of the data. Different structures of time are supported by appropriate views. In our tool, we used the linear time line structure, which can reveal trends and tendencies. Support of cyclic and calendric views is currently under development.

The kit was developed with a view that interactivity is key in visual knowledge discovery. The kit therefore supports brush, pen-and-zoom, and all mouse events out of the box. Our app makes use of the mouse events and programmatic interface to link the kit's time-series views with a map-based choropleth view (described below). Users can create a temporal alignment of event sequences to a common event or index date. Detection of patterns can be aided by performing temporal aggregation and noise reduction that can be applied or removed on the client side, without the need for resending of data, which can be time and bandwidth consuming.

Data Content

The tool receives location demarcation information in GeoJSON format, a geospatial data structure based on JavaScript Object Notation (JSON). In our demo, we use GeoJSON files representing Africa's countries and provinces, downloaded from Natural Earth (<http://www.naturalearthdata.com/>). Average temperature/rainfall of the period 1990 to 2012 was taken from The World Bank Climate Change Knowledge Portal (<http://sdwebx.worldbank.org/climateportal/>).

Generating Policy Dependent Simulations

Any framework can be used to drive the generation of scenario simulations for the tool, as long as it provides results that can be described as numeric variable values that are specific to policy, location, and time (in the required JSON format). Figure 1 describes a general framework for learning models from data and generating forecasts under multiple intervention policies. The models are fitted by statistical machine learning software using model specifications and data to produce learned (possibly location-specific) model parameters. The simulations are created by model simulators using fitted models in conjunction with existing data and specified intervention (policy) specifications. The simulations are then evaluated using evaluation modules that take the scenario simulation results and process them using the specifications of the evaluation measures and, possibly, cost specifications provided by modelers to produce predicted outcomes (objective values). The policy-planning-related results of the whole process are visualized in the decision support tool.

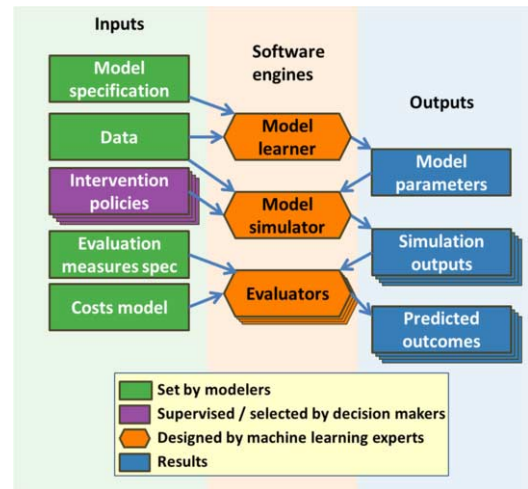


Figure 1 – Model learning and scenario simulation framework. Domain / location specific inputs are provided by modelers and policy planners in green and purple, respectively. These are processed using the appropriate software engines (orange), which may be general-purpose tools. Model learning and scenario simulation results are provided as outputs (blue).

To demonstrate the decision support tool, we constructed a weather-dependent generative malaria model. The exact details of the modeling work are not important to this report; they are not based on actual disease prevalence measurement data and therefore not meant to provide true estimates of malaria for the respective countries. Here is a high-level description to illustrate the complexity and richness of the dynamics involved:

We modeled the epidemic as a vector-host-weather model. State variables included temperature and precipitation (driven by annual mean + integrated fluctuations), Susceptible, Exposed, Infected and Recovered (SEIRS) proportions of humans in the population, and Susceptible, Exposed and Infected (SEI) number of mosquitos per human.

We modeled the dynamics by a set of Ordinary Differential Equations (ODEs), integrated using the Euler method with constant time step size and state dependent noise.

The model parameters that are determined by temperature and precipitation affect the dynamics of all other state variables included adult mosquito birth rate, adult mosquito per capita death rate, mosquito biting rate, proportion of bites by susceptible mosquitoes on infected humans that produce infection, duration of the sporogonic cycle, survival probability of infected mosquitoes over the incubation period of the parasite, and proportion of bites by infectious mosquitoes on susceptible humans that produce infection. A similar model was previously presented by Parham and Michael [10].

The modeled intervention options included the presence of Insecticide Treated bed-Nets (ITN) and Indoor Residual Spraying (IRS) for specified proportions of the human population. Their effects are somewhat different; ITN prevents bites but does not affect the size of the mosquito population, while IRS both prevents bites and slightly affects the mosquito death rates. ITN and IRS began at a value of zero in all policies and changed to a policy-determined fixed-value at the end of the

second year. Other patterns could have been applied (e.g., degrading effects).

Results

Policy Selection and Comparison

The first tab of the system displays the details of all available policies in separate panels (Figure 2). Each panel consists of the policy's title, a short description, and its cost and length parameters. The left pane enables policy filtering by the same parameters. The user may also select specific policies of interest and the tool will produce a comparison table for them. Except for policy title and description, the appearance of all other fields is configurable.

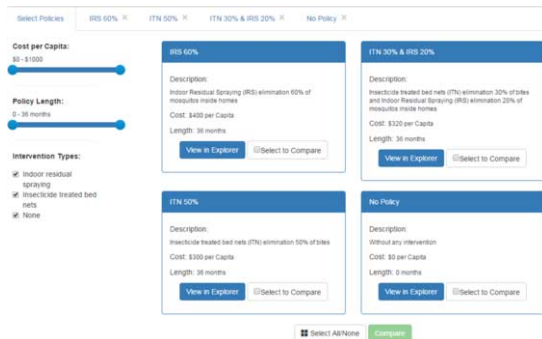


Figure 2 – Policies Selection tab. Each panel represents a single policy. Policies may be compared in a table or by opening those of interest in 'Scenario Explorer' tabs. The left pane enables policy filtering by cost, length, and intervention types.

Co-Representation of Spatial and Temporal Data

Each policy may be opened in a separate Scenario Explorer tab, where users can observe its effect in both time and space (Figure 3). The left pane is dedicated to visualizing temporal data for a selected country. The user may select which of the available charts should be visible and the desired time range. Each chart shows measurement (e.g., survey) data when available and model-generated estimates comprising the median value as well as a confidence interval. Hovering over a chart's line displays the median value and date at that time point; clicking on it modifies the display on the map to the right.

When a time point is clicked, the map's colors change to a choropleth, which shows the value of the selected chart and time point for each of the modeled regions. For example, the map in Figure 3 is colored by precipitation values on January 1, 2017. The selection of chart and time point also opens two legends: one shows the value range matching each color, and the other shows the values of all charts in this time point for the selected region. Hovering over different regions at this stage will change the latter to display the hovered region's current charts' values. Selecting a new region will load its time series data in the left pane, without changing the charts selected to be visible or the time range.

The user can add more policy Scenario Explorer tabs to present while preserving the configuration from previous ones. The tabs are synchronized so that selecting a region or a visible chart, setting the time range, or selecting a time point on a specific chart, updates all other policy tabs to show the same properties for their respective policies. As such, the user can make personalized configurations on one of the policy tabs,

and then easily compare the outcomes of other policies by simply switching between tabs.

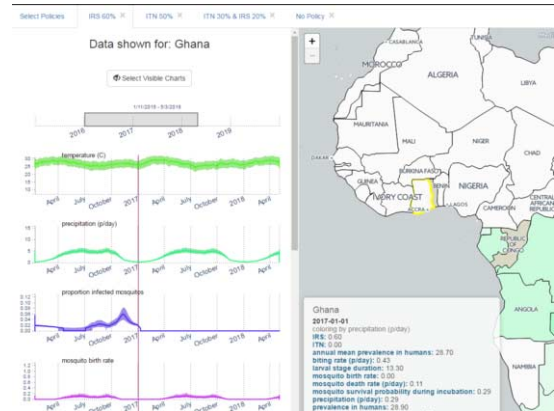


Figure 3 – Scenario Explorer tab. The left pane shows the data as time series charts for a specific country (Ghana). The right pane shows the data as a choropleth map of all African countries for a specific time point (01/01/17).

Proof of Concept: The Malaria Case

In order to demonstrate the tool's capabilities, we present a use case for malaria intervention in Africa. The model described in the methods was used to obtain malaria prevalence estimates for African countries for four ITN and IRS policies.

We obtained initial simulation values by running the simulation forward from a random initial condition for several annual cycles. The results differ across regions due to the differences in annual mean weather conditions, and across policies due to differences in the application and amount of IRS and ITN. For the purpose of this work, we generated four policies: with no intervention, with ITN at 30% and IRS at 20% from year 2, with ITN only at 50% from year 2, and with IRS only at 60% from year 2. For each location, 2000 simulation runs were conducted and the median, 25 and 75 percentiles were calculated per time point to extract inputs for the tool. Since malaria prevalence is seasonal, we extracted the mean prevalence in a running yearlong time window as a policy performance measure. The results of this measure are shown in Figure 4 for Ghana.

As can be seen, the mean prevalence differs across policies in the size of the median effect and in the variance of the predicted distribution. In the presented view, it is clear that the model predicts 60% IRS (Figure 4d) to be more effective in Ghana than the other policies, both in terms of reducing the median and in terms of minimizing variability of the measure. In a real policy decision support scenario, other performance measures such as policy costs would be taken into consideration in a multiple goal optimization setting. In this demo-scenario, the policy would present a tradeoff between policy effectiveness and costs shown in Figure 2, such that no one policy outperforms across both measures.

A Single Flexible Framework for Multiple Scenarios

The system can be easily adapted for multiple purposes, as it relies only on a few configuration files, and the data itself is held in structured files: The map polygons are generated from GeoJSON files (see Methods), meaning the countries display can be easily switched to display different geographical re-

gions (e.g., states, provinces or even houses) by providing the relevant set of files. Model results can be of any type supported by the time-oriented data visualization library and only need to match the geographical entities (see Methods). The policies are held in JSON files, as are the data by region and by time point for each policy. These files were uploaded to a NoSQL document-based database (see Methods), and as such have a flexible structure that can be easily edited or automatically generated (see Technical Appendix).

Thus, supporting different diseases, policies, or chart types, as well as results from entirely different domains is possible and requires minimal effort.

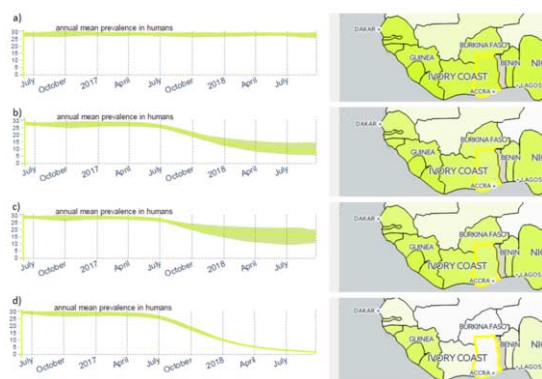


Figure 4 – Annual mean prevalence of malaria under four different types of interventions. The left pane shows the data for Ghana as a time series over the years 2016–2018. The right pane shows the data as a choropleth map for a single time point (05/20/2018) across different countries. Each row represents a different intervention policy: **a)** No policy implemented **b)** ITN for 30% of the population, IRS for 20% of the population **c)** ITN for 50% of the population, and **d)** IRS for 60% of the population.

Discussion

In this work, we present a web application that can assist healthcare policy decision makers in estimating the effect of different policies on different regions in the near and far future, allowing them to make informed decisions. The tool presents data in time series and across regions simultaneously for each policy, and allows for an easy comparison between policies. To the best of our knowledge, no other tool currently available provides these abilities in a consumable interface.

Many of the health systems of developing countries, as well as epidemiological studies and organizations, have reported using DHIS2 (<https://www.dhis2.org>) for data management, aggregation and display [14]–[16]. Indeed, DHIS2 offers an informative interface for the visualization and summarization of data gathered at multiple locations. It has analysis abilities over past data, enabling the generation of varied types of charts; however, our tool has two primary advantages in comparison.

First, DHIS2 does not offer inter-connectability between plots and maps, whereas our tool's map and charts are linked in ways that allow spatio-temporal data exploration. Future versions of the tool will also provide histogram views of selected variables and time points to reflect expected distribution across the population of the modeled regions.

Second, DHIS2 is not geared towards visualizing or comparing prediction data for multiple what-if scenarios, and does not measure or estimate distributions for a specific location and time. Thus, the ability to use this tool for future policy planning is limited. DHIS2 is, however, open source, modular, and highly extendable. Plans for future development include DHIS2 integration that will enable using our tool to visualize results and select policies, and then transfer selected policy prediction results back into DHIS2 so that they can be shared with the numerous users of that system.

Our tool was implemented in a manner that allows complete domain flexibility. It does not rely on specific types of data or regions. Any measurement and modeling results of a spatio-temporal nature, and any numeric calculated outcome measures can be visualized. Policies are defined by the types and values of intervention parameters and can be compared by any of the parameters or supplied (model generated) time dependent variables. The number of charts and their types can be easily configured, as well as the area of concern and its division into comparable regions.

With that said, the system still has some limitations. Its performance is linearly dependent on the number of charts, the number of time points, and the number of displayed policies. It is responsive to touch-screen interfaces (mobile devices), though the full map and charts view is better displayed on larger screens. Lastly, the actual modeling and the interface that runs the models to produce policy a prediction is currently separate from the tool. While this allows for greater flexibility in selecting modeling environments, future work will integrate the specification of policies and consequent running of models into the web-based interface.

Conclusions

We describe a novel tool for policy planning and decision support over spatial and temporal data. It offers visualization for the multi-dimensional effects of different policies, simultaneously across different regions and different time ranges. The tool, which focuses on user interaction and visualization, is, by design, separate from the modeling and simulation framework; this allows it to serve a wide range of models and modeling choices. Although we chose to demonstrate its abilities for the analysis of epidemiological data, it can be easily used for decision support in other spatio-temporal domains (e.g., agriculture). Models of infectious diseases and, especially, vector-borne diseases can be quite complex. Similarly, policy objectives can be numerous and their focus can span multiple time periods. Therefore, choosing the best policy involves not only accurate modeling, but also using tools that allow policy makers to view and explore complex statistical spatio-temporal information in a manner that is easy to digest.

Technical Appendix

Our database server is document-oriented. Each entity – in this case: policy, region, time point – is defined as a type within a JSON file, while all JSON files are held in the database and are stored in an optimized key-value paradigm. The JSON files formats divide into three types accordingly:

- Policy (Figure 5; top) – uniquely describes a single policy, including all the interventions implemented as a part of it

- Time point (Figure 5; middle) – contains the values of all regions on all charts in a single time point, under the policy defined in “policy_id”
- Region (Figure 5; bottom) – contains the values of all time points on all charts for a single region, under the policy in “policy_id”

```
sample_policy_description.json:
{
  "id": "<unique policy id>",
  "type": "policy",
  "title": "<title>",
  "description": "<policy description>",
  "cost": "<cost per capita>",
  "length": "<policy length in months>",
  "interventions": ["<intervention1>", "<intervention2>", ...],
  "data": [<any additional data>]
}

sample_policy_timepoint.json:
{
  "type": "timepoint",
  "id": "<unique time point>",
  "policy_id": "<policy id>",
  "data": {
    "<region 1>": {
      "<key for chart1>": "<value>",
      "<key for chart2>": "<value>",
      "<key for chart3>": "<value>",
      "<key for chart4>": "<value>",
      ...
    },
    "<region 2>": {
      "<key for chart1>": "<value>",
      "<key for chart2>": "<value>",
      "<key for chart3>": "<value>",
      "<key for chart4>": "<value>",
      ...
    },
    ...
  },
  ...
}

sample_policy_region.json:
{
  "type": "region",
  "id": "<unique region name>",
  "policy_id": "<policy id>",
  "data": [
    {
      "type": "<key for chart1>",
      "time": "<timepoint1>",
      "value": "<actual value>",
      "high": "<confidence interval max value>",
      "low": "<confidence interval min value>"
    },
    {
      "type": "<key for chart1>",
      "time": "<timepoint2>",
      "value": "<actual value>",
      "high": "<confidence interval max value>",
      "low": "<confidence interval min value>"
    },
    ...
    {
      "type": "<key for chartN>",
      "time": "<timepointM>",
      "value": "<actual value>",
      "high": "<confidence interval max value>",
      "low": "<confidence interval min value>"
    }
  ]
}
```

Figure 5 – Sample database documents

Acknowledgements

The work described here was conducted as part of collaboration between Notre Dame University and IBM Research.

References

- [1] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, *Nat. Phys.*, vol. 8, no. 1, pp. 32–39, Dec. 2011.
- [2] M. Marathe and A.K.S. Vullikanti, Computational epidemiology, *Commun. ACM*, vol. 56, no. 7, p. 88, Jul. 2013.
- [3] P. Klepac, C.J.E. Metcalf, A.R. McLean, and K. Hampson, Towards the endgame and beyond: complexities and challenges for the elimination of infectious diseases, *Philos. Trans. R. Soc. London B Biol. Sci.*, vol. 368, no. 1623, 2013.
- [4] S. Bhatt et al., The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015, *Nature*, vol. 526, no. 7572, pp. 207–211, Sep. 2015.
- [5] H. Heesterbeek et al., Modeling infectious disease dynamics in the complex landscape of global health, *Science*, vol. 347, no. 6227, 2015.
- [6] E. Odongo-Aginya, G. Ssegwanyi, P. Kategere, and P. C. Vuzi, Relationship between malaria infection intensity and rainfall pattern in Entebbe peninsula, Uganda, *Afr. Health Sci.*, vol. 5, no. 3, pp. 238–45, Sep. 2005.
- [7] M. Gambhir et al., Geographic and ecologic heterogeneity in elimination thresholds for the major vector-borne helminthic disease, lymphatic filariasis, *BMC Biol.*, vol. 8, no. 1, p. 22, 2010.
- [8] E. Michael and B.K. Singh, Heterogeneous dynamics, robustness/fragility trade-offs, and the eradication of the macroparasitic disease, lymphatic filariasis, *BMC Med.*, vol. 14, no. 1, p. 14, Dec. 2016.
- [9] E. Michael and R.C. Spear, Modelling Parasite Transmission and Control, Springer Science & Business Media, 2010.
- [10] P.E. Parham and E. Michael, Modeling the Effects of Weather and Climate Change on Malaria Transmission, *Environ. Health Perspect.*, vol. 118, no. 5, pp. 620–626, Dec. 2009.
- [11] P.E. Parham et al., Modeling the role of environmental variables on the population dynamics of the malaria vector *Anopheles gambiae sensu stricto*, *Malar. J.*, vol. 11, no. 1, p. 271, 2012.
- [12] S.A. Cushman and E.L. Landuth, Spurious correlations and inference in landscape genetics, *Mol. Ecol.*, vol. 19, no. 17, pp. 3592–3602, Sep. 2010.
- [13] P. Bak, A. Yaeli, I. Skarbovsky, and J. Bnayahu, Time Oriented Data Visualization Library for Data Scientists, in *Event: Proceedings of the IEEE VIS 2016 Workshop on Temporal & Sequential Event Analysis*, 2016.
- [14] M. Mera, C. González, and D.M. López, Towards an intelligent decision support system for public health surveillance - a qualitative analysis of information needs., *Stud. Health Technol. Inform.*, vol. 202, pp. 44–7, 2014.
- [15] V.M. Kiberu, J.K.B. Matovu, F. Makumbi, C. Kyozira, E. Mukooyo, and R. K. Wanyenze, Strengthening district-based health reporting through the district health management information software system: the Ugandan experience., *BMC Med. Inform. Decis. Mak.*, vol. 14, no. 1, p. 40, May 2014.
- [16] O. Akinwande et al., A Decade of Monitoring HIV Epidemics in Nigeria: Positioning for Post-2015 Agenda., *AIDS Behav.*, Nov. 2016.

Address for correspondence

Michal Chorev
IBM Research - Haifa, Israel
E-mail: michalc@il.ibm.com

The Pluripotent Rendering of Clinical Data for Precision Medicine

Christopher G. Chute^a, Stanley M. Huff^b

^a Chief Research Information Officer, Johns Hopkins Medicine, Baltimore, MD, USA,

^b Chief Medical Informatics Officer, Intermountain Healthcare, Murray, Utah, USA

Abstract

Health care and biomedical research are awash in data. Traditional data warehouse methodologies do not scale to this challenge; nor do their schema match the variety of analytic use cases. An alternative model, which shreds data into well-formed constituent data elements, conformant with the emerging CIMI-FHIR standards and stored together with the complete, raw, source data using modern and scalable data utilities such as Hadoop and its derivatives, affords the creation of pluripotent data repositories. Such repositories can be leveraged to generate any number of data marts, registries, and analytic data sets, each of which “just in time” binds an appropriate use-case specific data model. We call this notion PiCaRD: Pluripotent Clinical Repository of Data. We believe such nimble biomedical data management strategies are crucial for Precision Medicine discovery and application.

Keywords:

Archiving; Reference Standards; Common Data Elements.

Introduction -The Data Challenge

Health care and biomedical research have become profoundly information intensive [1]. It is inconceivable today to practice “best evidence” medicine without the monitoring of practice pathways, potential adverse events, and increasingly, genomic correlations. The depth and breadth of pertinent information long ago outstripped the capacity of humans to know and understand the full spectrum of pertinent knowledge, never mind to accurately and reliably recall all that should be considered. From an analytic perspective, we are seeing the aggregation of data, such as health claims at national scale, with the integration of data arising from electronic health records (EHRs) and their supporting systems, such as detailed departmental sources, diagnostic devices, ‘omics repositories, PACTS, and ICU waveforms. Perhaps most significantly, it has become self-evident that within five years, a far greater amount of clinically relevant data about our patients will be generated outside of the healthcare delivery industry than within it. Specifically, the proliferation of smart-phone and home monitoring devices, tied wirelessly into personal health networks, is starting to generate continuous waveform monitoring of pulse, blood pressure, respiratory rate, blood glucose, patient activity, location, ambient light, other environmental exposures, and sleep quality, to name but a few examples. Image data is capturing detailed diet information, risk behaviors, occupational hazards, and living conditions.

Traditional methods for managing these tsunami of detailed, heterogeneous, and complex data will not scale to the

emergent challenges of discovery research, population health management, continuous quality improvement, comparative effectiveness analyses, patient centered outcomes research or operations research. We propose a re-thinking of how health systems regard health-related data, curate it for inferencing and discovery, and engage in information partnerships with their populations and patients.

Precision Medicine

The advent of Precision Medicine poses additional demands on data access. Based on the principle that detailed knowledge of a patient can precisely inform optimal treatments and interventions, the idea relies on access to vast clinical information resources that can inform these decisions. During the process of assembling such data resources, the problems of large data volume, heterogeneous semantics and syntax, and dissonant data schema all present challenges.

Several countries have engaged in large-scale population studies, collecting rich datasets of many data types including clinical, genomic, and mobile-device generated data. The United States, through its Precision Medicine Initiative [2], now rebranded as the “All of Us” project [3], will generate datasets that exemplify the challenges outlined above. We are posing our solution to address these and related datasets that can enable facile analytics for Precision Medicine.

Traditional Data Repositories

Most health systems consider the information about their patients to be contained in their administrative systems and their EHRs. Yet healthcare delivery in the US is palpably fragmented, with information about patients distributed among many EHRs for any given patient to the extent that it measurably distorts phenotyping and analytics [4]. Internally, departmental and diagnostic systems generate huge volumes of data, only the veneer of which typically is transferred to EHRs and which is almost always stripped of its metadata and provenance. For example, ECG tracings are typically reduced to static images with reporting parameters, generated by stand-alone machines. ICUs and floor monitors generate highly detailed waveform data, which are either discarded outright or collapsed into quantitative snapshots such as blood pressure, oxygen saturation, temperature, and related physiological measurements. Again, the untapped potential in these waveform data offers demonstrable contributions to our understanding and management of many disease manifestations [5].

Most clinical data warehouses are a derivative of the EHR, and thus suffer from all the lossy transforms and metadata pruning that occurs. While such data reduction may be

judicious for the acute management of most patients, the information loss cannot but compromise deeper data analyses, association finding, and metric development to deliver better care. Finally, while Meaningful Use [6] is mitigating the data comparability and consistency challenge in the United States, multicenter research and networks such as CTSA and PCORI still confront the reality that most data repositories reflect the idiosyncrasies of localized and divergent coding systems, data models, and renderings that seem to remain the birthright of every healthcare delivery organization. Much of this data can be and often is translated into a canonical representation, though the perennial question of what constitutes the canonical information standard is debated across interest groups, standards development organizations, and analysts.

Relational Systems

The pioneering work of Edgar Codd on relational algebra and set theory in the 1960s and the subsequent development of relational databases [7] has dominated the database world for decades. Most database implementations invoke the relational model and the by now truly standard Standard Query Language (SQL). These are enormously efficient data structures to answer questions when the query design is known; however, when queries imply a different schema, they can be unworkably non-performant. For example, most EHRs organize data by patient, which means that a query across all patients can be inefficient, if not unmanageable. Clinical and translational research, by their nature, encounter questions that cannot be anticipated in any given data structure. For research retrieval many have stated, “all schema are wrong.” For this reason, alternatives to SQL data structures are increasingly emergent for research queries, though clinical data repositories have not yet widely adopted those structures.

Scalable Data Structures

The emergence of Not Only SQL databases (NOSQL), such as Mongo and Couch DB, have relaxed the constraints of relational schema with only minimal impact on retrieval efficiency. Most such databases are document-centric, where the information is stored as an annotated document, aligning with an XML data definition or schema. An alternative is the entity-attribute-value structure, manifest in the MUMPS architecture as perhaps one of the original NOSQL data structures.

The advent in the 1980s of big-data management environments, orthogonal to the nearly coincident emergence of practical relational databases, has spawned what many regard as a notable, fundamental shift in computing, generally classified under Stonebraker’s influential “shared nothing architectures” [8] and initially popularized by companies such as Tandem and Teradata. This architectural design supports nearly infinite scaling, because the independent and self-sufficient component parts of the system “share nothing,” such as file systems or memory, with the others. Apache Hadoop [9] is a relatively recent open-source project implementing “share nothing” design, thus allowing for the distributed processing of large datasets across clusters of computers. The suitability of Hadoop and its derivatives for clinical and -omic data has been previously described [10].

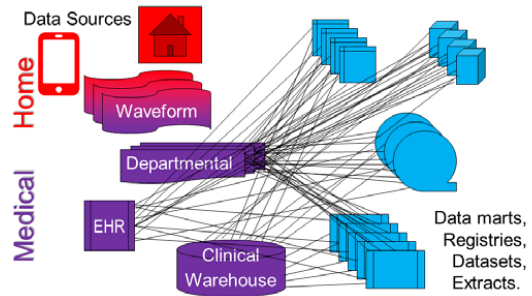


Figure 1- This figure depicts the traditional “point to point” connections between data sources and target data marts, registries, and analytic datasets. For each of these connections, a separate, frequently duplicative, data normalization and harmonization process may be defined, designed, implemented, and maintained. Clinical warehouses may not capture all relevant data and metadata, and thus investigators frequently resort to source systems in departments and the EHR. In this diagram, waveforms and data sources from the home and mobile devices are not part of the data extraction “ecosystem;” this is not uncommon in medical center infrastructures.

Self-describing Data Objects

The tension between chaotic data heaps (no schema) and constraints on data expression and retrieval (relational schema) suggests an optimal point in the middle, a just-right Goldilocks level of schema. Clinical Information Modeling Initiative (CIMI) [11] data element objects are designed to be that optimal level of structure, ensuring data coherency and interpretability without imposing premature binding on data relationships and structure. Nevertheless, the opportunity to create relational or other schema, a late-binding approach, is preserved through metadata and provenance information carried along with the clinical data element object. CIMI data elements can be regarded as a more fully specified rendering of FHIR [12] resources, with which CIMI maintains close alliance.

The payload of any data object may be misinterpreted if the context and provenance of the data collection are missing. Thus crucial to clinical data element interpretation and filtering is the metadata, including date times, workflow context, and key relationships to other data objects (such as a drug-level test explicitly related to a drug order). Additionally, a cache of the raw, untransformed data should always be stored as part of the data element object, to enable retransformation or correction of the normalized CIMI-based rendering.

Pluripotent Data Repository

What we are proposing and promoting is a “next generation” clinical data repository framework, premised on a collection of pluripotent clinical data elements, which we call PiCaRD: Pluripotent Clinical Repository of Data. These data elements are managed in modern “share nothing” data architecture.

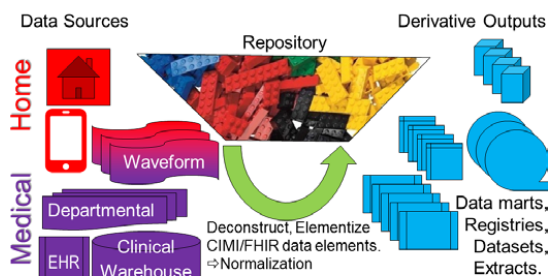


Figure 2- The pluripotent repository captures complete data from all sources, deconstructing that data into its constituent elements, rendered as data elements using FHIR/CIMI “micro-schema.” These are depicted as Lego pieces, managed and indexed by the Hadoop d Hadoop derivative, Accumulo environment. An arbitrary number of derivative data marts, registries, and analytic datasets can be extracted and maintained by the system; the normalization into a canonical form (FHIR/CIMI data elements) is done just one time. Each derivative resource manifests a use-case specific data schema or relational model such as i2b2, OMOP, OHDSI, or the PCORI CDM, illustrating late binding. The normalized repository comprises a hub, in a hub and spoke configuration, contrasting with the point-to-point configuration in Figure 1.

While the management environment is arbitrary, it must be able to accommodate a large number of complex data objects, scale to big-data performance and storage requirements, preserve content integrity and security, and efficiently support query and extraction. Our preliminary experimentation at the Johns Hopkins University has leveraged Apache Accumulo [13], which is an open-source Hadoop derivative, enhanced by the National Security Agency to include cell-level encryption and authentication, and fast content indexing (Figure 2).

Data Normalization

Anticipating that a health system will accumulate data from a variety of heterogeneous sources, and aspire to interoperate in federated programs and projects with other organizations, the data element payloads must be comparable and consistent internally and externally; this implies rendering the data payload in a canonical form. Which standard rendering we wish to declare canonical is frequently fractious; however, the CIMI-FHIR partnership appears to be gaining momentum, adoption, and utility. The SHARPN consortium [14-15], led by two of us (CC, SH), produced a large body of open-source normalization pipelines [16] designed to render heterogeneous clinical data elements; albeit the canonical formalism was the Clinical Element Model [17], a predecessor to the current CIMI-FHIR framework. We are re-working this software to produce CIMI-specific data elements and become production-grade resources.

The real elegance of a standards-based, pluripotent soup of data elements is that information may be integrated, normalized, and managed from an arbitrary number of sources. In SHARPN we demonstrated, for example, that drug data from HL7 messages, EHR medication lists, pharmacy dispensing, pharmacy benefits managers, and mentions in free text could all be rendered into a common formalism, although with computably recognizable differences in provenance metadata. Thus a health system can meaningfully integrate rich, original data from departmental systems, home monitoring, personal devices, EHR transactions, and claims

data into a common framework. The careful preservation of complete, raw source data in parallel protects against incomplete preservation of data in any particular canonical form, and secular change in the normalized specification. Correspondingly, that framework would sustain federated inquiry across different organizations, assuming appropriate consent and authorization were in place.

Derivative Data Marts, Registries, and Analytic Datasets

The resulting “data soup” (a data lake full of self-describing data objects), replete with bulky data elements, is likely to be a suboptimal environment for analytic inferencing and report generation. However, such a pluripotent repository can generate an arbitrary number of derivative data marts, registries, and analytic datasets on demand or periodically. These derivative data collections can be bound to an appropriate data schema or relational model, where iterative reporting and inferencing can proceed efficiently. This concept of “late binding” to a data schema optimized for the data-collection use case is critically important. For example, one could generate an arbitrary number or variations of i2b2, OMOP/OHDSI, PCORI Common Data Model, or other common research data models from a common instance of a PiCaRD repository. More interestingly, data derivatives that are highly specific to a single protocol or analytic use case can be generated as easily.

Precision Medicine Application

The recent announcement that the US Precision Medicine Initiative (PMI) will use FHIR interfaces to enable patients to forward clinical information to the coordinating center is a strong advocacy for using FHIR structures for PMI discovery research. Specifically the *Sync for Science* tooling [18] toolkit being developed for the harvesting of clinical data by patients in the PMI is based upon the FHIR specification. Over time, this will enormously simplify data upload into a PiCaRD repository, as many of the normalizations steps will be unnecessary in the face of native FHIR data object transfer from *Sync for Science*.

Conclusion

Historical data warehouse designs, while addressing many practical needs, are unlikely to scale to the scopes demanded by health systems in this information-intensive age. A more nimble framework, invoking data schema and structure only at the level of a clinical data element, can demonstrably capture an unbounded number of raw and processed data types. Using current “share nothing” data models such as Hadoop derivatives, it can also perform effectively in big data problems. Such a resource can in turn generate an arbitrary number of data marts, registries, and analytic data sets, each of which can manifest use-case specific optimizations of schema and data models. This pluripotent data framework will not displace traditional data warehouses but can be the foundation on which purpose-specific data warehouses can be efficiently generated and maintained.

Acknowledgements

This work was supported in part by 90TR000201-4 *SHARP Area 4: Secondary Use of EHR Data* from the US Office of the National Coordinator in the Department of Health and Human Services, USA.

References

- [1] D. DeMets, L. Tabak, R. Altman, D. Botstein, A. Califano, D. Ginsburg, P. Hurn, D. Masys, J. Mesirov, S. Murphy, and L. Ohno-Machado, Data and Informatics Working Group Report to The Advisory Committee to the Director, in, NIH, Bethesda, 2012, pp. 1-92.
- [2] F.S. Collins and H. Varmus, A new initiative on precision medicine, *N Engl J Med* **372** (2015), 793-795.
- [3] National Institutes of Health: NIH, PMI Cohort Program announces new name: the All of Us Research Program, in, 2016.
- [4] W.Q. Wei, C.L. Leibson, J.E. Ransom, A.N. Kho, P.J. Caraballo, H.S. Chai, B.P. Yawn, J.A. Pacheco, and C.G. Chute, Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus, *J Am Med Inform Assoc* **19** (2012), 219-224.
- [5] F.M. Kashif, G.C. Verghese, V. Novak, M. Czosnyka, and T. Heldt, Model-based noninvasive estimation of intracranial pressure from cerebral blood flow velocity and arterial pressure, *Sci Transl Med* **4** (2012), 129ra144.
- [6] D. Blumenthal and M. Tavenner, The "meaningful use" regulation for electronic health records, *N Engl J Med* **363** (2010), 501-504.
- [7] E.F. Codd, A Relational Model of Data for Large Shared Data Banks, *Communications of the Acm* **13** (1970), 377-387.
- [8] M. Stonebraker, The Case for Shared Nothing, *Database Engineering* **9** (1986), 4-9.
- [9] The Apache Software Foundation, Apache™ Hadoop®, in, 2015.
- [10] C.G. Chute, M. Ullman-Cullere, G.M. Wood, S.M. Lin, M. He, and J. Pathak, Some experiences and opportunities for big data in translational research, *Genet Med* **15** (2013), 802-809.
- [11] Clinical Information Modeling Initiative, CIMI, in, 2015.
- [12] HL7, Fast Healthcare Interoperability Resources (FHIR), in, 2015.
- [13] The Apache Software Foundation, Apache™ Accumulo™, in, 2015.
- [14] C.G. Chute, J. Pathak, G.K. Savova, K.R. Bailey, M.I. Schor, L.A. Hart, C.E. Beebe, and S.M. Huff, The SHARPN project on secondary use of Electronic Medical Record data: progress, plans, and possibilities, *AMIA Annu Symp Proc* **2011** (2011), 248-256.
- [15] S. Rea, J. Pathak, G. Savova, T.A. Oniki, L. Westberg, C.E. Beebe, C. Tao, C.G. Parker, P.J. Haug, S.M. Huff, and C.G. Chute, Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPN project, *J Biomed Inform* **45** (2012), 763-771.
- [16] SHARPN Consortium, Strategic Health IT Advanced Research Projects (SHARP): Secondary Use of EHR Data, in.
- [17] Intermountain Healthcare, Clinical Element Browser, in.
- [18] J. Mandel, Sync for Science: Helping patients share EHR data with research, in, 2016.

Address for correspondence

Christopher G. Chute, MD
 Bloomberg Distinguished Professor of Health Informatics
 Professor of Medicine, Public Health, and Nursing
 Johns Hopkins University
 2024 E. Monument Street
 Suite 1-202
 Baltimore, Maryland 21205
chute@jhu.edu

Classification of Clinical Research Study Eligibility Criteria to Support Multi-Stage Cohort Identification Using Clinical Data Repositories

James J. Cimino,^a William J. Lancaster,^b and Mathew C. Wyatt^a

^a Informatics Institute, University of Alabama at Birmingham, Birmingham, Alabama, USA

^b Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA

Abstract

One of the challenges to using electronic health record (EHR) repositories for research is the difficulty mapping study subject eligibility criteria to the query capabilities of the repository. We sought to characterize criteria as “easy” (searchable in a typical repository), “hard” (requiring manual review of the record data), and “impossible” (not typically available in EHR repositories). We obtained 292 criteria from 20 studies available from ClinicalTrials.gov and rated them according to our three types, plus a fourth “mixed” type. We had good agreement among three independent reviewers and chose 274 criteria that were characterized by single types for further analysis. The resulting analysis showed typical features of criteria that do and don’t map to repositories. We propose that these features be used to guide researchers in specifying eligibility criteria to improve development of enrollment workflow, including the definition of EHR repository queries for self-service or analyst-mediated retrievals.

Keywords: Informatics, Biomedical Research, Electronic Health Records

Introduction

While paper-based patient health records have a long history of being used to identify potential research study subjects, the advent of electronic health records (EHRs) has greatly facilitated researchers’ access to relevant patient data, especially when they are transferred to data repositories or warehouses specifically designed for this purpose.[1] Researchers may use such information for estimating the availability of eligible research subjects in some larger target population (cohort estimation), identifying specific patients who might be interested in enrolling in a research study (cohort identification), or whose data themselves (whether in summary or detailed form) can be used to explore research questions (data reuse).[2] What most such endeavors have in common is some sort of study-specific set of patient characteristics, known as *eligibility criteria*, that must be present for a patient to be included as a study subject (*inclusion criteria*) or, if present, render the patient unsuitable for inclusion in the study (*exclusion criteria*).[3] Use of EHR data for subject identification will typically involve a repository query to find patients meeting inclusion criteria, eliminating patients based on exclusion criteria.

Ideally, a researcher should be able to access a repository in “self-service” mode to carry out the search and retrieval processes independently. However, the required queries are often too complex for this approach and, in such cases, require the

assistance of a data analyst who is familiar with the data and tools and can mediate between the researcher and the repository.[3,4] Even then, the process may become complicated when the eligibility criteria are expressed in some complex arrangement that is not amenable to automated searching or when they require data that are either not accessible through the query tools or not present in the health record.[5] Studies of the use of EHR repositories to match enrollment criteria have had success rates ranging 23%[6] to 44%[5] due to requirements for temporal restrictions or calculated conditions, or because inaccuracies in medical diagnostic coding.

For example, a study might seek patients with diabetes mellitus who have been receiving escalating doses of insulin and then arrive in an emergency room with an elevated blood glucose level. The researcher might be able to identify patients with diabetes who have received insulin but can only determine whether they meet the escalating dose criteria through manual analysis of the records. The use of the repository fails when identifying patients in an emergency room with an acute problem, since the repository data are retrospective. Subject recruitment is further complicated because not only must the potential subjects be identified immediately, they must be enrolled prior to treatment in the acute setting. Researchers with such complex, time-sensitive eligibility criteria must often resort to educating the front-line clinicians (or posting reminders) and hoping that they will be notified in time to enroll subjects. Automated alerts, tied to the EHR can be used if the eligibility criteria are simple,[7] but will be difficult to implement in complex situations with which even analysts struggle.

One solution, which is often carried out in practice, mimics the use of a screening protocol in which a large set of potentially eligible patients is identified and each individual record is examined for specific criteria. The EHR repository version of this method involves the researcher or analyst carrying out an initial query to identify patient records that meet some criteria and then reviewing full records manually for the rest, perhaps flagging patients who might be eligible if they appear in care settings for further evaluation.

We believe that this multi-step process can be facilitated by characterizing eligibility criteria in advance as either being amenable to retrieval with repository tools (“Easy”) or requiring some more elaborate mechanism (“Hard”). If a researcher specifies criteria in this way in their research protocols, the task of the repository user can then be reduced to a feasible retrieval step to identify an initial cohort that is then passed on to the next stage in the process for application of additional criteria through other means. The purpose of the study presented here was to explore the characterization of eligibility criteria to determine if such an approach is feasible.

- 1 – “Easy” – The criterion could be used easily to identify subjects in an EHR data repository, assuming the data were present, using the repository’s user interface
- 2 – “Mixed” – The criterion could be partly retrieved using the repository’s user interface but then would require further manual review
- 3 – “Hard” – Matching the criterion would require manual review of the record
- 4 – “Impossible” – The repository would not be likely to include sufficient information to match patients to the criterion, either because the source EHR would not be likely to contain the information or the repository would not include the information in a timely manner

Figure 1– Rating Scale for Eligibility Criteria. Each criterion was assessed by each rater for its suitability for retrieval from an electronic health record data repository.

Methods

We obtained the Clinical Trials.gov (NCT) identifiers (a US registry of clinical trials that includes enrollment criteria) for a convenience sample of 10 studies conducted at UAB that were chosen for a separate study of the use of EHR data for cohort prediction (unpublished data). We then incremented each NCT ID successively *until* it matched an additional study. For example, the NCT ID NCT01098981 would be incremented to NCT01098994 to find the additional study. These twenty studies provided eligibility criteria that served as the data set for our study. The NCT record for each study was examined to identify eligibility criteria (see, for example, <https://clinicaltrials.gov/ct2/show/NCT01098981>) which were copied from a Web browser screen to a spreadsheet.

The authors of this study served as experts for rating the eligibility criteria. JJC was the principle architect of the National Institutes of Health’s Biomedical Translational Research Information System (BTRIS)[8] and has extensive experience with its use for matching patients to research criteria.[9-12] WJL was a medical resident University of Alabama School of Medicine (UASOM) at the time of the study, with extensive experience with electronic health records and moderate experience with UASOM’s i2b2 (Informatics for Integrating Biology to the Bedside)[13] repository. MCW is a systems analyst architect, with responsibility for, and experience using both i2b2 and PowerInsight (Cerner Corporation, Kansas City, MO), the clinical data warehouse attached to the EHR used by UASOM.

Each of the authors reviewed the resulting criteria and rated them with a scale (shown in Figure 1) that was intended to capture the rater’s judgement as to whether a typical clinical data repository, such as i2b2, BTRIS or PowerInsight, would be able to identify subjects matching the criterion based on information available in a typical EHR. Raters were permitted to assign multiple scores in cases where a “single” criterion actually included multiple clinical concepts. The ratings were compiled and summarized to determine the characteristics of criteria that could be readily used with EHR repositories, versus those that could not.

Results

Criteria Data Set: A sample of the clinical trials used in this study are listed in Table 1. The cleaning process yielded a set of 301 criteria, of which 292 were found to be exactly or essentially unique each of which was rated by each rater (876 ratings in all). Examples are shown in Table 2.

Criteria Ratings: In 29 of the 876 ratings (3.3%), involving 18 of the 292 criteria (6%), reviewers found some difficulty as-

signing a single code to a complex criterion (see Table 2). When these 18 criteria were excluded from further analysis, all three raters agreed on the same rating 140 times for the remaining 274 criteria (50.9%), with least two of three raters agreeing 256 times (93.1%). Three raters agreed or one rater differed from the other two raters by one (for example, two said “1”, and the third said “2”) on 244 (89.1%) of the criteria. Agreement between pairs of raters ranged from 66.5 to 66.9%, with Pearson’s Correlations between .7901 and .8178 ($p < .00001$ for all comparisons). Table 2 shows examples of ratings and Tables 3 and 4 summarize the interrater statistics.

Characteristics of Easy Criteria: Raters had unanimous or nearly unanimous agreement that 80 criteria were “easy” (with complete agreement in 85% of cases). In general, these were related to the categories of data that are provided by repositories like i2b2 and BTRIS, such as laboratory results, vital signs, diagnoses, procedures and allergies. The raters were undaunted by criteria that included temporal restrictions such as “digoxin within 6 months of starting treatment” and “parathyroidectomy within 12 weeks prior to screening”. In addition, criteria involving complex phenotypes were considered “easy” if they could be addressed through an assemblage of individual easy criteria.

Characteristics of Mixed Criteria: Raters had agreement or near agreement that 70 criteria were “mixed” (with 47% in complete agreement), that is, the criterion could be partly retrieved using the repository’s user interface but then would require further manual review. These criteria generally included some mention of an “easy” criterion, coupled with some restriction or co-occurring state or procedure that either was not itself an easy criterion or the relationship between the two could only be carried out by a manual review.

Table 1. Research descriptions from Clinical Trials.gov that were selected for this study. Studies were paired to include ten UAB studies and ten studies that followed them sequentially in the Clinical Trials.gov database.

NCT ID	URL	Study Title
010989 94	Haptoglobin Phenotype, Vitamin E and High-density Lipoprotein (HDL) Function in Type 1 Diabetes	https://clinicaltrials.gov/ct2/show/NCT01098994
013822 12	A Study to Evaluate the Safety of Paricalcitol Capsules in Subjects With Chronic Kidney Disease	https://clinicaltrials.gov/ct2/show/NCT01382212
013822 25	Sodium Hyaluronate Ophthalmic Solution, 0.18% for Treatment of Dry Eye	https://clinicaltrials.gov/ct2/show/NCT01382225
017974 45	Study to Evaluate the Safety and Efficacy of E/C/F/TAF (Genvoya®) Versus E/C/F/TDF (Stribild®) in HIV-1 Positive, Antiretroviral Treatment-Naive Adults	https://clinicaltrials.gov/ct2/show/NCT01797445

Table 2. Examples of criteria ratings. A, B and C correspond to co-authors; 1-4 correspond to categories from Figure 1.

* indicates criteria for which there was complete or near agreement, included in further analysis.

A	B	C	Criterion
1	1	1	Abnormal liver function as defined in the protocol at Screening *
1	1	1	absolute neutrophil count $\geq 1,000/\text{mm}^3$ *
1	1	1	allergy to starch powder or iodine. *
1	1	2	digoxin within 6 months of starting treatment. *
1	1	1,2	Anti-diabetic drug naïve, or treated with one or two oral antidiabetic drug (OADs), or treated with human Neutral Protamin Hagedorn (NPH) insulin or long-acting insulin analogue or pre-mixed insulin, both types of insulin either alone or in combination with one or two OADs
1	4	x	For women, effective contraception during the trial and a negative pregnancy test (urine) before enrollment
2	2	2	Active or untreated latent tuberculosis (TB) *
2	2	4	Women of childbearing potential that are pregnant, intend to become pregnant, or are lactating
2	3	4	Dermal disorder including infection at anticipated treatment sites in either axilla.
3	3	3	HDSS score of 3 or 4. *
3	4	4	Able to swallow tablets *
3	4	4	good overall physical constitution *
3	4	4	Legal incapacity or limited legal capacity *
4	4	4	Ability to understand and sign a written informed consent form, which must be obtained prior to initiation of study procedures *
4	4	4	Facial hair *
4	4	4	Willing to be examined *
4	1,4	4	Current alcohol or substance use judged by the investigator to potentially interfere with study compliance

Characteristics of Hard Criteria: Raters had agreement or near agreement that 50 criteria were “hard” (with 70% in complete agreement), that is, matching the criterion would require manual review of the record. These criteria span the breadth of EHR findings that typically appear in physicians’ notes and procedure reports. Some were parts of “mixed” criteria, such as “females who are breastfeeding”. Others are statements of patient condition that would not be reflected in a typical problem list, such as “prisoner”. Scale-based assessments, such as “Pre-morbid modified Rankin score of 0-1” were common. Finally, some criteria might require the researcher to use the clinical record to make the assessment, such as “Candidate for phototherapy or systemic therapy”.

Characteristics of Impossible Criteria: Raters had agreement or near agreement that 44 criteria were “impossible”, that is, the repository would not be likely to include sufficient information to match patients to the criterion. Complete agreement occurred in only 24 cases (55%). While some of the “impossible” criteria related to the time factor of the data (such as “Parents/children who refuse to participate in the study” and “presenting within timeframe for intravenous tPA treatment approved by local regulatory authorities but no more than 4.5 hours from onset of symptoms”), other cases depended on raters’ opinions about which relevant data simply do not find

their way into EHRs, such as “Facial Hair”, “good health”, and “Able to swallow tablets”.

Table 3. Summary of results of criteria rating.

Criteria	
Total	301
Unique	292
Three single ratings	274 (93.8%)
All in Agreement	140 (51.1%)
At Least Two Raters in Agreement	256 (93.4%)
All in Agreement or Two in Agreement & Third Off by One	244 (89.1%)

Table 4. Inter-rater comparisons. A, B and C are raters (co-authors). 1-4 are categories from Figure 1.

	A&B	B&C	A&C
Interrater Agreement	183 (66.8%)	191 (69.7%)	183 (66.8%)
Pearson’s Coefficient	0.8077	0.8178	0.7901
P	<.00001	<.00001	<.00001

General Recommendations: The data presented above can be summarized by separating out “easy”, “hard” and “impossible” criteria into explicit lists (see Figure 2). This list, in turn, can be used by those developing eligibility criteria such that they can anticipate the workflow needed to identify potential research subjects using repository query tools, manual record review, and real-time interventions at recruitment sites.

Criterion is likely to be easy to retrieve from an EHR repository if:

- they correspond to simple data types (e.g., laboratory, medications, diagnoses, procedures, vital signs)
- they have simple temporal restrictions (start or stop dates, but not before or after other events)
- phenotypes that can be completely described by the above criteria

Criterion is likely to be hard to retrieve from an EHR repository if:

- a restriction is placed on a simple criteria, such as severity, location or treatment status of a condition.
- a calculation based on simple criteria
- the finding is reported as a rating on a standardized scale it requires information that would be likely to appear in the text of a clinician’s note or procedure report
- a researcher’s judgement is needed regarding suitability, based on information in the record

Criterion is likely to be impossible to retrieve from an EHR repository if:

- it is not typically captured in EHRs
- it must occur close to the time of enrollment
- it can only be determined during enrollment (such as willingness to undertake or forego some activity)

Figure 2 - Features that suggest they will be easy, hard or impossible to retrieve from a typical EHR

Discussion

The use of EHR data for research is a rapidly evolving phenomenon. The development of methods for “large pragmatic trials” is, in particular, reliant on EHRs.[14] Yet the methods for defining eligibility criteria remain largely unchanged in practice, with little apparent attention paid to the fact that the criteria elicited from a patient filling out a questionnaire might differ in fundamental ways from those elicited from an EHR.

A great deal of informatics research has examined the semantics of clinical research eligibility criteria. Projects such as the Ontology of Clinical Research (OCRe),[15] and the Agreement on Standardized Protocol Inclusion Requirements for Eligibility (ASPIRE)[16] have defined broad classes of criteria (such as demographics, disease-specific features, functional status, etc.) with the intent of making them computable for comparison across studies. The semantic complexity of these classes of criteria has been studied in depth by Ross and colleagues.[17] Weng and colleagues provide a comprehensive review of this work[18] and extended it with detailed analyses of the *semantic types* of criteria,[19] including the use of the National Library of Medicine’s Unified Medical Language System[20] and NIH’s Common Data Elements.[21] To our knowledge, however, there have been no systematic studies of how such criteria translate operationally to what can and can’t be accomplished with EHR repository query tools.[2-4]

We did not set out to create a reusable criteria rating scale, but rather sought to simply find a specific set of criteria that could be used for further analysis. Nevertheless, we experienced fairly good interrater correlation in the use of our scale. Although this study was carried out by one set of researchers at one institution, we are of diverse backgrounds (medicine, informatics, and information systems) and draw on experience with three very different repositories (Power Insight, i2b2 and BTRIS). Examples of criteria on which we agreed and disagreed are included in Table 2 for the reader to judge whether our rating method appears generalizable.

Our examination of a randomly collected set of enrollment criteria suggests that a large percentage (at least 89%) can be readily classified as easy, hard or impossible to retrieve from EHR data repositories. We do not make any claims, given our sample size and sampling method, about the actual ratios between easy, hard and impossible criteria across the spectrum of clinical trials. But, it seems logical to take advantage of the knowledge that such distinctions exist for developing the methods to be used for matching patients to criteria when EHR data repositories are being used for at least a part of the process. It should not be hard to imagine an interactive “wizard” that queries researchers for criteria in structured ways that ask first for easy elements that can be addressed with a repository search tool:

- What laboratory test result would you like and with what range of values?
- What class of medications should the patient be taking (or not taking) and for what duration?
- What procedure should have already been, or not been, performed?
- Are diagnoses from billing data sufficient or are higher-quality sources (such as problem lists) required?

It might then proceed to the more complex elements that require manual review of records retrieved with the initial criteria:

- What calculation would you like performed on the previously specified laboratory test result?
- What should the indication for the previously specified medication be?
- What findings should be present in, or absent from, the text report for the previously requested procedure?
- What complex phenotypic pattern should the patient have?

It could then move on to elements that can only be ascertained at the time of enrollment:

- What acute event will the otherwise eligible patient have (or not have) at the time of presentation?
- What will the patient have to agree to do (or not do) in order to be eligible for enrollment?

This hypothetical wizard might then produce a coherent set of eligibility criteria that could specify a multi-staged enrollment process that starts with a set of data queries, proceeds to manual review of full records, and then specifies what additional information will be needed on specific potential subjects through direct contact, perhaps by flagging them in an EHR’s alerting system to notify the researcher when they appear in a patient care setting.[7]

The findings of our study are only a building block in the development of an approach to improve the specification of eligibility criteria. An understanding of the pragmatics of interfacing with actual clinical data repositories can inform previous work on the semantics and interoperability of criteria. A set of questions, such as those listed above, guided by knowledge of criteria semantics, would help to establish a dialogue between the investigator and the data analyst (or provide an opportunity for introspection if the investigator is using a repository in a self-service mode). Further improvement would likely require an iterative process of extensions, corrections and refinements.

Conclusions

The findings of this study support the hypothesis that clinical research eligibility criteria fall into stereotypical categories with respect to data available in EHR repositories. By describing these categories explicitly, the possibility exists that a structured process, perhaps using a data entry form or interactive “wizard”, could be developed that would allow researchers to express their criteria in ways that would allow establishing realistic expectations for the recruitment process and improve retrieval of data, either by analysts or the researcher herself.

Acknowledgements

This work was supported by research funds from the Informatics Institute of the University Of Alabama School of Medicine and by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR001417. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- [1] H. Cao, M. Markatou, G.B. Melton, M.F. Chiang, and G. Hripcsak, Mining a clinical data warehouse to discover disease-finding associations using co-occurrence statistics, *AMIA Annu Symp Proc.* 2005:106-10.
- [2] I. Danciu, J.D. Cowan, M. Basford, X. Wang, A. Saip, S. Osgood, J. Shirey-Rice, J. Kirby, and P.A. Harris, Secondary use of clinical data: the Vanderbilt approach, *J Biomed Inform.* 2014 Dec;52:28-35.
- [3] H. J. Lowe, T.A. Ferris, P.M. Hernandez, and S.C. Weber, STRIDE--An integrated standards-based translational research informatics platform, *AMIA Annu Symp Proc.* 2009 Nov 14;2009:391-5.
- [4] J.J. Cimino, E.J. Ayres, A. Beri, R. Freedman, E. Oberholtzer, and S. Rath, Developing a self-service query interface for re-using de-identified electronic health record data, *Stud Health Technol Inform.* 2013;192:632-6.
- [5] V.G. Deshmukh, S.M. Meystre, and J.A. Mitchell, Evaluating the informatics for integrating biology and the bedside system for clinical research, *BMC Med Res Methodol* **9** (2009), 70.
- [6] E.K. Johnson, S. Broder-Fingert, P. Tanpowpong, J. Bickel, J.R. Lightdale, and C.P. Nelson, Use of the i2b2 research query tool to conduct a matched case-control clinical research study: advantages, disadvantages and methodological considerations, *BMC Med Res Methodol* **14** (2014), 16.
- [7] R.A. Jenders, G. Hripcsak, R.V. Sideli, W. DuMouchel, H. Zhang, J.J. Cimino, S.B. Johnson, E.H. Sherman, and P.D. Clayton, Medical decision support: experience with implementing the Arden Syntax at the Columbia-Presbyterian Medical Center, *Proc Annu Symp Comput Appl Med Care* (1995), 169-73.
- [8] J.J. Cimino, E.J. Ayres, L. Remennik, S. Rath, R. Freedman, A. Beri, Y. Chen, and V. Huser, The National Institutes of Health's Biomedical Translational Research Information System (BTRIS): design, contents, functionality and experience to date, *J Biomed Inform* **52** (2014), 11-27.
- [9] J.J. Cimino, Normalization of phenotypic data from a clinical data warehouse: case study of heterogeneous blood type data with surprising results, *Stud Health Technol Inform* **216** (2015), 559-63.
- [10] J.D. Manning, B.E. Marciano, and J.J. Cimino, Visualizing the data - using Lifelines2 to gain insights from data drawn from a clinical data repository, *AMIA Jt Summits Transl Sci Proc* (2013), 168-72.
- [11] J.J. Cimino, The false security of blind dates: chrononymization's lack of impact on data privacy of laboratory data, *Appl Clin Inform* **3**(4) (2012), 392-403.
- [12] J.J. Cimino, L. Farnum, G.E. DiPatrizio, and B.R. Goldspiel, Improving adherence to research protocol drug exclusions using a clinical alerting system, *AMIA Annu Symp Proc* (2011), 257-66.
- [13] S.N. Murphy, M. Mendis, K. Hackett, R. Kuttan, W. Pan, L.C. Phillips, V. Gainer, D. Berkowicz, J.P. Glaser, I. Kohane, and H.C. Chueh, Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside, *AMIA Annu Symp Proc* (2007), 548-52.
- [14] J.D. Lurie and T.S. Morgan, Pros and cons of pragmatic clinical trials, *J Comp Eff Res* **2**(1) (2013), 53-8.
- [15] I. Sim, S.W. Tu, S. Carini, H.P. Lehmann, B.H. Pollock, M. Peleg, and K.M. Wittkowski, The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research, *J Biomed Inform* **52** (2014), 78-91.
- [16] I. Sim I and J.C. Niland, Chapter 9: Study protocol representation, in R.L. Richesson and J.E. Andrews, editors, *Clinical Research Informatics*, Springer-Verlag; London, (2012), 172.
- [17] J. Ross, S. Tu, S. Carini, and I. Sim, Analysis of eligibility criteria complexity in clinical trials, *AMIA Summits Transl Sci Proc* (2010), 46-50.
- [18] C. Weng, S.W. Tu, Sim I, and R. Richesson, Formal representation of eligibility criteria: a literature review, *J Biomed Inform* **43**(3) (2010), 451-67.
- [19] Z. Luo, M. Yetisgen-Yildiz, and C. Weng, Dynamic categorization of clinical research eligibility criteria by hierarchical clustering, *J Biomed Inform* **44**(6) (2011), 927-35.
- [20] C. Weng, X. Wu, Z. Luo, M.R. Boland, D. Theodoratos, and S.B. Johnson, EliXR: an approach to eligibility criteria extraction and representation, *J Am Med Inform Assoc* **18** (suppl) (2011), i116-24.
- [21] Z. Luo, R. Miotto, and C. Weng, A human-computer collaborative approach to identifying common data elements in clinical trial eligibility criteria, *J Biomed Inform* **46**(1) (2013), 33-9.

Address for correspondence

James J. Cimino, M.D ciminoj@uab.edu

Translational Morphosyntax: Distribution of Negation in Clinical Records and Biomedical Journal Articles

K. Bretonnel Cohen^{a,c}, Foster R. Goss^b, Pierre Zweigenbaum^c, Lawrence E. Hunter^a

^a Computational Bioscience Program, University of Colorado School of Medicine, Aurora, CO, USA

^b University of Colorado School of Medicine, Department of Emergency Medicine, Aurora, CO, USA

^c LIMSI, CNRS, Université Paris-Saclay, Orsay, France

Abstract

Prior knowledge of the distributional characteristics of linguistic phenomena can be useful for a variety of language processing tasks. This paper describes the distribution of negation in two types of biomedical texts: scientific journal articles and progress notes. Two types of negation are examined: explicit negation at the syntactic level and affixal negation at the sub-word level. The data show that the distribution of negation is significantly different in the two document types, with explicit negation more frequent in the clinical documents than in the scientific publications and affixal negation more frequent in the journal articles at the type level and token levels. All code is available on GitHub¹.

Keywords:

Natural Language Processing; Data Mining; Linguistics

Introduction

Natural language processing (the treatment of human language by computers) is increasingly used in health care and in biomedical research [1–4]. Negation is a frequent cause of errors in language processing [5–9]. This has stimulated a considerable amount of work on negation in clinical documents (e.g. the classic work cited above and recent work summarized in [10]), and a lesser amount of work on negation in scientific literature (e.g. the work summarized in [11–14], as well as scientific literature corpus creation efforts [15–19]). However, while that work has made considerable progress, it has focused almost exclusively on explicit negation by words with phrasal scope, such as *no* and *not*. In contrast, negation at a sub-word level—what has been called *affixal* or *morphological negation*, such as the *de-* in *dephosphorylate* or the *a-* in *afebrile*, meaning *without fever*—has received very little attention in biomedical language processing. This is a gap in the literature, as this kind of negation has implications for many things in biomedical language processing and biomedical communication in general, ranging from lexicon/terminology design, to readability of and access to health care information by non-specialists, to the performance of natural language processing applications. Furthermore, the majority of work on negation in the biomedical domain has focused on evaluation of system performance; very little of it has looked at distributional characteristics of negation in the relevant genres. This is a serious gap because it has implications for our understanding of the system performance that is the topic of most work on negation in the domain. This

paper addresses both of those issues. In particular, we look at both clinical data and scientific literature and compare them with respect to their distributions of two kinds of negation: explicit negation (words such as *no* and *not*), and morphological negation (the *a-* in *afebrile*). Along the way we discuss a data set that we have prepared containing several thousand ambiguous words marked as to whether or not they begin with a negative prefix. The null hypotheses that we evaluate are that there are no differences in the distribution of negation between clinical and scientific biomedical texts at any of the levels to be examined; as will be shown, in fact there are such differences at the syntactic level and at the morphological level, and the differences are sometimes large.

This paper takes a distributional and descriptive approach because in text mining and natural language processing, knowledge of the distribution of any linguistic phenomenon can help us predict the contribution of that phenomenon to error rates in our applications. Distribution of negation in particular is important both in natural language processing and in language science more generally. In a paper that we discuss in more detail below, [20] Wu et al. point out that distributional characteristics of linguistic phenomena can have deep implications for evaluating not just individual systems, but also for evaluating the literature on a topic overall, where the performance that is reported in a paper may accurately describe the performance of a system when it is optimized for a specific dataset, but not be generalizable. This can lead to the conclusion that a particular problem is essentially solved, when in fact all that has been solved is dealing with a particular definition of that problem in a specific data set. In particular, Wu et al. point out that there has been considerable work on negation in biomedical text, particularly in clinical text, with a smaller body of work existing on negation in scientific journal articles. They describe a number of published solutions to negation in clinical texts, as well as observe that they are optimized to particular genres of text, and that those solutions do not necessarily generalize well at all.

From the point of view of system development and evaluation, knowledge of the distribution of a linguistic phenomenon can help select suitable document sets to use for some specific task type (at the granularity of, say, parsing, coreference resolution, etc.). It can help prioritize module development, and in the case of negation, it may interact with usability of tools based on natural language processing, given what we know about human processing of negation from psycholinguistic studies: negated assertions are more difficult to process [21]. Distributional information can be used as a form of prior

¹<https://github.com/KevinBretonnelCohen/NegationDistribution>

knowledge in machine learning applications that allow supervised under- and over-sampling [22].

Finally, to the authors' knowledge, there is very little literature that explores translational issues in negation. All work that we are aware of in the biomedical domain has looked either at clinical texts, or at published biomedical literature. To address that gap in the literature, we have compared two very different genres with relevance to translational research: on the one hand, scientific journal articles, and on the other hand, clinical documents.

Context of the present work

There has been a small amount of previous work on the distribution of negation. Yaeger-Dror and Tottie [23] focused on spoken versus written English and found differences in the distribution of affixal and explicit negation when comparing spoken versus written language. Chapman et al. [6] examined the distribution of explicit negation *within* a genre and found that it may be Zipfian, noting that "The negation algorithm was triggered by sixty negation phrases with just seven of the phrases accounting for 90% of the negations." Subsequent work found this to be true across multiple Germanic languages, as well as the Romance language French [24]. Cohen et al. [25] compared the abstracts and bodies of full-text journal articles and found that article bodies had a higher percentage of explicit negatives, at 5.3/thousand words versus 3.8/thousand words in abstracts ($p < .01$ by Mann-Whitney-Wilcoxon), concluding that this was relevant to the relative difficulty of information extraction from the two text types. Verspoor et al. [26] made the same measurements in Open Access versus traditional journals, finding no significant difference between them and using that to argue that Open Access journals are representative of the biomedical literature as a whole. Kjellmer studied some of the interesting problems of affixal negation of adjectives in English, such as which adjectives can and cannot be negated (e.g., English has *unkind*, but not *uncruel*), and includes data on distribution of affixal negatives across types of adjectives [27]. Globally, this body of work can be summarized as showing that the distribution of negation is structured, and that it can be shown to vary in interesting ways (or not) both within and across genres. However, it remains the case that the literature on differences between morphological and explicit negation is very small, to the point that there have not been opportunities to evaluate the replicability of the associated findings, and the topic has not been addressed at all in the biomedical domain.

Wu et al. [20] point out that one of the consequences of the sublanguage nature of clinical documents is that there is a limited number of ways to express negation; this is true, but previous studies of negation in clinical literature have focused on negation at the syntactic level. Here we extend the domain of inquiry into a previously unstudied part of the grammar of biomedical text: the morphological level. Wu et al. pointed to the morphological differences in annotations as a possible explanatory factor that was uncharacterized. The work discussed here adds a considerable amount of data to that discussion, adding the ability to compare clinical data (the subject of the small amount of previous observations about distribution of negation in the biomedical domain) with data on scientific publications. Using the same processing on both data sources makes the results directly comparable, which has not been the case with previous work.

Methods

Materials

Since the goals of this study are translational in nature, the materials for this work were drawn from the clinical domain and from the biomedical literature: *MIMIC II progress notes* on the one hand [28,29] and the *CRAFT corpus* on the other [30–32]. The rationale behind this choice of domains is that they are close to the opposite ends of the spectrum between bench (the mouse being a common model organism) and the bedside. Other choices could potentially be useful, e.g. journal articles with a clinical orientation, other kinds of clinical documents, etc.

MIMIC II progress notes: Half a million words of physician-written progress notes from the MIMIC II corpus. They reflect the status of patients in the Intensive Care Unit.

CRAFT corpus: A corpus of scientific journal articles in the domain of mouse genomics, previously shown to be representative of the biomedical scientific literature [26].

Explicit negation

We took samples of 10,000 consecutive words from both document types, for a total of 440,000 words each (the closest total sample size to the number of words in CRAFT, the smaller corpus). We counted the number of explicit negative words per 10,000-word sample. The set of explicit negatives that we counted was: *no|not|none|denies|nothing*. Details of the normalization can be found in the script, available on GitHub and named *negativesEvery10KWords.pl*. (One could argue about the completeness of the set of explicit negatives that was used in the experiment, but any omissions would affect both text types proportionally and would not be likely to change the overall conclusions of the study.)

Affixal negation

All word types in both corpora were collected, and after normalizing for case and punctuation, the number of tokens of each type was counted. To clarify the meanings of the terms *type* and *token*: the word *denaturation* occurs 9 times in the CRAFT corpus. The word *ativan* appears 22 times in our sample of the MIMIC II corpus. These represent two *types* (*denaturation* and *ativan*), and a total of 31 *tokens*. (See script *directoryToTypeTokenCounts.pl* on GitHub for details of the normalization.) Then we extracted all words beginning with any string that can be a negative prefix in English. To ensure objectivity, we obtained definitions of the set of negative prefixes in English from neutral third parties, named in a file on GitHub. This step produced a list of 5,196 words that can be thought of as ambiguous with respect to whether or not they begin with a negative morpheme. The extracted words were examined manually and classified as actually containing a negative prefix, or not. All words from both document sets were presented as single words in isolation.

With the judgements about which words did and did not begin with a negative prefix, along with the counts of each of those words, we calculated the total number of tokens beginning with an actual negative prefix in each document collection.

To build the set of words beginning with ambiguous strings, we first searched the two text collections for words beginning with the following character sequences, all of which are listed as negative prefixes in the 3rd-party sources listed on the GitHub site: *un, no, a, de, dis, anti, il, im, in, and ir*.

The resulting set of words can be thought of as ambiguous with respect to whether or not they begin with a negative morpheme, since it contains words such as *antiapoptotic* (CRAFT, 6 tokens), *immature* (MIMIC, 177 tokens), and *desaturation* (MIMIC, 72 tokens), which begin with negative morphemes, and *anticipated* (CRAFT, 3 tokens), *improved* (MIMIC, 205 tokens), and *detailed* (MIMIC, 809 tokens), which do not. These are the words which were manually classified as beginning with a negative morpheme, or not.

Guidelines

We developed the guidelines in three rounds, consisting of

1. a test of an initial set of guidelines on Amazon Mechanical Turk²,
2. a subsequent test of a revision of the guidelines on Amazon Mechanical Turk,
3. ...and then review of the guidelines by all authors.

In developing the Mechanical Turk tasks, we took note of the ethical guidelines for crowdsourced linguistic data in [33].

Even after those three rounds of guideline development our domain-expert annotators—one of whom had written the guidelines—still had questions about some specific cases. The final set of guidelines is available on GitHub.

To ensure the clarity and consistency of the guidelines and the neutrality of the resulting counts, the data was double-annotated. The annotators represent typical readers of the materials in question: an emergency room physician and a former registered cardiovascular technologist with a PhD in linguistics and a specialty in biomedical language.

For the final annotation step, the word types from CRAFT and MIMIC II were combined into a single file and randomized (both with respect to corpus and with respect to ranking within each corpus). The inter-annotator agreement was 0.94 before resolution, and the entire calculation of agreement is documented on GitHub.

Finally, we used a two-tailed t-test to assess the statistical significance of the observed differences in explicit negation, and the chi square test for the data on affixal negation.

Replicability, repeatability, and reproducibility

All scripts, annotation guidelines for the affix study, and judgments of the individual annotators are on GitHub (see URL at the bottom of the first page). The annotation project and analysis were repeated by an independent third party to ensure that it was replicable. CRAFT is available at bionlp.sourceforge.net. MIMIC II requires a data use agreement, but is freely available.

Results

Explicit negation

The distribution of explicit negatives for the two document collections is shown in Figure 1. The distributions are quite

different, with a mean of 111 per 10,000-word sample for the MIMIC II progress notes, and a mean of 31 per 10,000-word sample for the CRAFT corpus. A Welch 2-sample t-test shows a statistically significant difference, $t = -27.092$, $df = 53.822$, $p\text{-value} < 2.2e-16$.

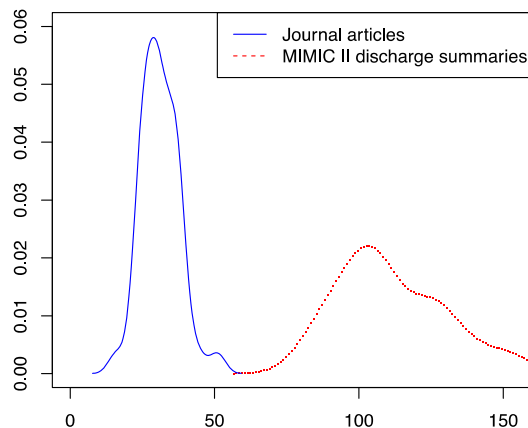


Figure 1 - Density distribution of frequency of explicit negation per 10,000 words, MIMIC II and the CRAFT corpus.

Affixal negation

The distribution of morphologically negated and non-negated words is shown in Table 1, along with their ratios on the type level and on the token level. The distribution of morphological negation is different at the type level in the two genres, chi square = 8866.8, $df = 1$, $p\text{-value} < 2.2e-16$, with the journal articles having a higher incidence of morphologically negated types (0.028) than the clinical documents (0.017). Additionally, the distribution of morphological negation is also different at the token level in the two genres, chi square = 14338, $df = 2$, $p\text{-value} < 2.2e-16$, with the journal articles having a higher incidence of morphologically negated tokens, although the magnitude of the difference is smaller than that at the level of types (0.013 for CRAFT, 0.012 for MIMIC II).

Table 1 - Counts and ratios of negated types and tokens.

Corpora and counts	CRAFT	MIMIC
Negated types	650	319
Ambiguous non-negated types	2,641	1,586
Non-ambiguous non-negated types	19,545	16,400
Ratio of negated types to non-negated types	0.028	0.017
Negated tokens	5,575	6,763
Ambiguous non-negated tokens	63,819	70,945
Non-ambiguous non-negated tokens	367,576	84,526
Ratio of negated tokens to non-negated tokens	0.013	0.012

Discussion

The results of the hypothesis tests can be summarized thus: the distributions of explicit negation are different between the two genres by two-tailed t-test, with the clinical notes having a much higher incidence of negation than the journal articles. The distribution of affixal negation is different at the type level in the two genres by chi square, with the journal articles having a higher incidence of morphologically negated types. The distribution of affixal negation is also different at the

² Words from MIMIC II used in the AMT tests of the guidelines were manually screened by an author with HIPAA and human subjects training to ensure that those words did not contain any identifying or potentially identifying information. This was in addition to the screening that has already been done by the MIMIC Consortium.

token level in the two genres by chi square, with the journal articles having a higher incidence of morphologically negated tokens, although the magnitude of the difference is smaller than that at the level of types. Thus, there are differences in the distribution of negation between the two genres at both levels: explicit negation and morphological negation. Furthermore, the difference is in different directions at the two levels. At the level of explicit negation, there is more negation in the clinical texts. In contrast, at the affixal level, there is more negation in the journal articles.

These findings are especially relevant to a translational perspective on biomedical natural language processing, since the experiments reported here compared scientific literature to a clinical textual genre. From that perspective, the implications of the findings are that language processing systems that target the mapping of findings from the scientific literature to electronic health records will need to take a nuanced approach to handling negation, taking into account the different distributional characteristics of negational phenomena in the two genres.

Conclusions

In a 2014 paper with far-reaching implications for the study of negation, and indeed for machine learning in natural language processing in general, Wu et al. [20] concluded from their analysis of generalization versus optimization in clinical-domain negation detection systems that the best way to improve performance in negation detection is to manually annotate more data. In particular, they refer not to increasing the sizes of the corpora that we already have, but to annotating negation in data drawn from other distributions besides the corpora that are already available. The work reported here is a contribution in that direction, as one of the results of the work is a large set of words from clinical records and scientific journal articles, available at the GitHub repository, that have been annotated for the presence of a derivational, prefixal negation morpheme. Because the methodology that we describe here can yield relatively rapid judgements with good inter-annotator agreement, this two-corpus study can be rapidly extended to additional scientific and clinical genres.

In addition to the relevance of these findings to biomedical language processing, there are also implications for the construction of semantic resources for the domain. The community's investment in lexical, terminological, and ontological resources continues to be strong. The findings that we report here have implications for the approach to building those resources. Modern lexical-semantic resources such as PropBank and VerbNet [34–36] include separate entries for predicates that are related by the negative prefixes that have been studied in this paper. The Open Biomedical Ontologies seem to be following this strategy. However, since they have large numbers of “reversible” state-changing predicates, they do not seem to be keeping up, and if they can, may find the explosion in the number of terms to be overwhelming. For example, the Gene Ontology currently (file go-basic, version releases/2016-12-24) contains 8 terms that begin with *phosphorylation* (up from 6 in 2014), but only has 3 of the corresponding terms beginning with *dephosphorylation* (unchanged from 2014). A mechanism for dealing procedurally with this kind of prefixation could considerably reduce the maintenance load of biomedical resources like the Open Biomedical Ontologies. (Van Son et al. [37] gives an idea of what an affixal negation resource might look like, demonstrating the feasibility of the necessary annotation

tasks.) Thus, there are many potential applications for the distributionals that are reported on here.

Reproducibility

The code and data necessary to repeat /replicate this analysis are available for download at the GitHub repository named at the bottom of the first page of the paper. Future work that would be potentially revealing in terms of the reproducibility of the results reported here include at least the following, some of which are variations on the approach and some of which sample different populations:

- Using different sources of negation patterns, such as the most recent set of NegEx patterns, or those that could be mined from negation-annotated corpora, such as BioScope [12] and the BioNLP-ST shared task corpora [11,14].
- Expanding from the MIMIC II physicians' notes to other types of clinical data, from the CRAFT corpus to other scientific domains, and from English to other languages.

Acknowledgements

The work reported here was supported by NIH grants LM008111 and LM009254 to Lawrence E. Hunter, and NSF grant IIS-1207592 to Lawrence E. Hunter and Barbara Grimpe. Cohen's work was supported by Hunter's grants, by grant AHRQ R21HS024541-01 to Foster Goss (as was Goss's), and by generous funding from Labex DigiCosme (project ANR11LABEX0045 DIGICOSME), operated by ANR as part of the program Investissement d'Avenir Idex ParisSaclay (ANR11 IDEX000302), as well as by a Jean d'Alembert fellowship from the Fondation Campus Paris-Saclay as part of the « Investissement d'Avenir » program operated by ANR. PZ received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant 676207.

The authors thank Tiffany Callahan, who repeated the analysis to ensure that it was replicable, finding a bug in the process. We assure the excellent MEDINFO reviewers that the next paper will be titled *Of mice and men*, as suggested. We also thank Stephen Wu, Wendy Chapman, and the members of the group *Information Langue Écrite et Signée of the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Centre National de la Recherche Scientifique* for discussion. Mayla Boguslav, Harrison Pielke-Lombardo, Jingbo Xia, and Elizabeth White gave helpful comments on the manuscript.

References

1. Névél A, Zweigenbaum P. Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare. *Yearb Med Inform* (2015) **10**:194–8. doi:10.15265/IY-2015-035
2. Névél A, Zweigenbaum P. Clinical Natural Language Processing in 2015: Leveraging the Variety of Texts of Clinical Interest. *IMIA Yearb* (2016) 234–239. doi:10.15265/IY-2016-049
3. Simpson MS, Demner-Fushman D. “Biomedical Text Mining: A Survey of Recent Progress,” in *Mining Text Data* (Boston, MA: Springer US), 465–517. doi:10.1007/978-1-4614-3223-4_14
4. Velupillai S, Mowery D, South BR, Kvist M, Dalianis H. Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis. *Yearb Med Inform* (2015) **10**:183–93. doi:10.15265/IY-2015-009
5. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan B. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp* 2001 (2001) 105–109.
6. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A

- simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* (2001) **34**:301–310.
7. Goldin I, Chapman WW. Learning to detect negation with “not” in medical texts. *Proc ACM SIGIR* (2003)
 8. Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents. *J Am Med Informatics Assoc* (2001) **8**:598–609.
 9. Grouin C, Deléger L, Rosier A, Temal L, Dameron O, Van Hille P, Burgun A, Zweigenbaum P. Automatic computation of CHA2DS2-VASc score: information extraction from clinical texts for thromboembolism risk assessment. *AMIA. Annu Symp proceedings AMIA Symp* (2011) **2011**:501–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22195104> [Accessed December 24, 2016]
 10. Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc* (2011) **18**:552–556. doi:10.1136/amiajnl-2011-000203
 11. Kim J-D, Ohta T, Pyysalo S, Kano Y, Tsujii J. Overview of BioNLP’09 shared task on event extraction. *Proc Nat Lang Process Biomed NAACL 2009 Work Shar task* (2009)
 12. Vincze V, Szarvas G, Farkas R, Mora G, Csirik J. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* (2008) **9**.
 13. Morante R, Liekens A, Daelemans W. Learning the scope of negation in biomedical texts. *Proc Conf Empir Methods Nat Lang Process* (2008) 715–724.
 14. Nédellec C, Bossy R, Kim J-D, Kim JJ, Ohta T. Overview of BioNLP shared task 2013. in *BioNLP Shared Task 2013 Workshop*, 1–7.
 15. Sanchez-Grailliet O, Poesio M. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics* (2007) **23**:i424–32. doi:10.1093/bioinformatics/btm184
 16. Thompson P, Iqbal SA, McNaught J, Ananiadou S. Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinformatics* (2009) **10**:349.
 17. Pyysalo S, Ginter F, Heimonen J, Bjorne J, Boberg J, Jarvinen J, Salakoski T. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics* (2007) **8**:50.
 18. Wilbur JW, Rzhetsky A, Shatkay H. New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics* (2006) **7**.
 19. Agarwal S, Yu H, Kohane I. BioNOT: a searchable database of biomedical negated sentences. *BMC Bioinformatics* (2011) **12**:420. doi:10.1186/1471-2105-12-420
 20. Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C, Chapman W, Bridewell W, Hanbury P, et al. Negation’s Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing. *PLoS One* (2014) **9**:e112774. doi:10.1371/journal.pone.0112774
 21. Larrivé P, Lee C. *Negation and polarity: Experimental perspectives*. Springer International Publishing (2015).
 22. Cruz Díaz N, Maña López MJ, Mata Vázquez J, Pachón Álvarez V. A Machine-Learning Approach to Negation and Speculation Detection in Clinical Texts. *J Am Soc Inf Sci Technol* (2012) **63**:1398–1410.
 23. Yaeger-Dror M, Tottie G. Negation in English speech and writing: A study in variation. (1993).
 24. Chapman WW, Hillert D, Velupillai S, Kvist M, Skeppstedt M, Chapman BE, Conway M, Tharp M, Mowery DL, Deleger L. Extending the NegEx lexicon for multiple languages. in *Studies in Health Technology and Informatics*, 677–681.
 25. Cohen KB, Johnson HL, Verspoor K, Roeder C, Hunter LE. The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics* (2010) **11**:492. doi:10.1186/1471-2105-11-492
 26. Verspoor K, Cohen KB, Hunter L. The textual characteristics of traditional and Open Access scientific journals are similar. *BMC Bioinformatics* (2009) **10**:183. Available at: <http://www.biomedcentral.com/1471-2105/10/183>
 27. Kjellmer G. Negated Adjectives in Modern English: A corpus-based study. *Stud Neophilol* (2005) **77**:156–170.
 28. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K SH. PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. *Circulation* (2000) **101**:215–220.
 29. Saeed M, Villarreal M, Reisner AT, Clifford G, Lehman L-W, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit Care Med* (2011) **39**:952–960.
 30. Bada M, Eckert M, Garcia K, Evans D, Sitnikov D, Baumgartner WA, Ogren P V, Lanfranchi A, Howard A, Corvey W, et al. The Colorado Richly Annotated Full-Text (CRAFT) Corpus: A Resource for BioNLP Research. *Proc Seventh Rocky Mt Bioinforma Conf* (2009) 35.
 31. Verspoor K, Cohen KB, Lanfranchi A, Warner C, Johnson HL, Roeder C, Choi JD, Funk C, Malenkiy Y, Eckert M, et al. A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. *BMC Bioinformatics* (2012) **13**:
 32. Cohen KB, Verspoor K, Fort K, Funk C, Bada M, Palmer M, Hunter LE. “The {Colorado Richly Annotated Full Text (CRAFT)} Corpus: Multi-Model Annotation In The Biomedical Domain,” in *Handbook of Linguistic Annotation*, eds. N. Ide, J. Pustejovsky (Springer).
 33. Sagot B, Fort K, Adda G, Mariani J, Lang B. Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé. in *TALN’2011 - Traitement Automatique des Langues Naturelles* Available at: <http://hal.inria.fr/inria-00617067>
 34. Palmer M, Kingsbury P, Gildea D. The {P}roposition {B}ank: an annotated corpus of semantic roles. *Comput Linguist* (2005) **31**:71–106.
 35. Kipper K, Korhonen A, Ryant N, Palmer M. A large-scale classification of English verbs. *Lang Resour Eval* (2008) **42**:21–40.
 36. Palmer M, Gildea D, Xue N. *Semantic role labeling*. Morgan and Claypool (2010).
 37. van Son CM, van Miltenburg CWJ, Morante-Vallejo R. Building a Dictionary of Affixal Negations.

Address for correspondence:

Kevin Bretonnel Cohen, kevin.cohen@gmail.com

Extracting Sexual Trauma Mentions from Electronic Medical Notes Using Natural Language Processing

Guy Divita^{a,b}, Emily Brignone^{a,c}, Marjorie E. Carter^{a,b}, Ying Suo^{a,b},
Rebecca K. Blais^{a,c}, Matthew H. Samore^{a,b}, Jamison D. Fargo^{a,c}, Adi V. Gundlapalli^{a,b}

^a VA Salt Lake City Health Care System, Salt Lake City, Utah, USA

^b University of Utah School of Medicine, Salt Lake City, Utah, USA

^c Utah State University Department of Psychology, Logan, Utah, USA

Abstract

Patient history of sexual trauma is of clinical relevance to healthcare providers as survivors face adverse health-related outcomes. This paper describes a method for identifying mentions of sexual trauma within the free text of electronic medical notes. A natural language processing pipeline for information extraction was developed and scaled to handle a large corpus of electronic medical notes used for this study from US Veterans Health Administration medical facilities. The tool was used to identify sexual trauma mentions and create snippets around every asserted mention based on a domain-specific lexicon developed for this purpose. All snippets were evaluated by trained human reviewers. An overall positive predictive value (PPV) of 0.90 for identifying sexual trauma mentions from the free text and a PPV of 0.71 at the patient level are reported. The metrics are superior for records from female patients.

Keywords:

Natural Language Processing; Information Retrieval; Trauma and Stressor Related Disorders

Introduction

Patient history of sexual trauma is of great clinical relevance to healthcare providers. Sexual trauma is associated with a wide range of adverse health-related outcomes. These include psychiatric comorbidities such as PTSD, depressive disorders, anxiety disorders, and alcohol and drug related disorders [1-4], as well as medical comorbidities such as chronic pulmonary disease, liver disease, and hypertension [1-5]. Adverse social determinants of health include sexual re-victimization, homelessness, and suicide [6-9]. The initial psychiatric sequelae of sexual trauma, in the absence of intervention, may accumulate to affect a broad range of health and psychosocial domains, with symptoms often persisting for several years following the incident. Despite these clear implications for health status and health care needs, sexual trauma is generally poorly documented in healthcare settings [10].

The documentation of sexual trauma in administrative data such as International Classification of Diseases (ICD) coding is poor and likely under-represents the true prevalence of this complex problem. Furthermore, documentation of a history of sexual trauma in the free text of medical notes is also limited. While this may be a preference of the patient or practice of providers, the end result may be decreased access to appropriate trauma-informed care for survivors of sexual trauma [11].

The free text of the electronic medical note presents an opportunity to extract additional details pertaining to exposure to sexual trauma. Patients have the opportunity to disclose or express details of an experience of sexual trauma of any kind to a provider during a healthcare encounter. In addition, long term clinical follow-up allows the patient time to become comfortable with the provider and clinic, which may further facilitate disclosure.

This paper describes a method for identifying positively asserted mentions of sexual trauma within the free text of VHA electronic medical notes. Sexual trauma mentions are identified by a dictionary lookup algorithm using a curated dictionary of terms related to sexual trauma. The method also identifies the assertion status of the trauma mentions. The performance metrics, challenges, and relevant note types identified in this study will inform future NLP applications. Our novel application of NLP has broad generalizability in that it can be extended to other healthcare settings, as references to sexual trauma are likely universal.

As opposed to private and public health care systems in the US, Veteran's Health Administration (VHA) represents an integrated health care setting in which history of sexual trauma is regularly identified and documented. In 2004, VHA initiated universal screening of all Veterans seeking care in its medical facilities for exposure to sexual trauma during military service (military sexual trauma, MST). Currently, over 90% of new VHA users are screened, and 22% of women and 1% of men screened report having experienced this type of sexual trauma [12]. Under this model, many thousands of patients with a history of sexual trauma have been identified, facilitating access to specialized care designed to mitigate adverse outcomes. However, sexual trauma is underreported, and several estimates suggest that the prevalence of MST is much higher than official estimates based on structured screening [13, 14]. Additionally, screening focuses solely on sexual trauma occurring in a military setting, leaving survivors of childhood sexual trauma or non-military sexual trauma at risk of unmet health needs.

With respect to the VHA electronic medical record, a history of MST may be documented in several ways. Structured data include the result of a screen for MST described above. In the free text of electronic medical notes, there are a variety of techniques used by providers to document the results of screening questionnaires for MST. These include MST status indications written using grammatically well-formed sentences, but more frequently they appear in boiler-plated check box, question-answer, or slot value format. In this, special attention had to be given to identifying MST status boiler-plated mentions. It was necessary to further examine such boiler-plated mentions to insure that those

mentions were a true MST status indication. A number of boiler-plated mentions, which are in question-answer form are missing the answer. This includes missing indications in check box and slot-value formats as well. There is a further confounding of the boiler-plated format wherein the check boxes and slot-values are null; in this situation, it is not clear whether the provider has administered the MST survey with a checklist of items to discuss or whether the answer is null.

V3NLP Framework

The work described in this paper is built using the V3NLP Framework [15]. This framework provides a set of functionalities designed for building NLP applications that combine one reader and one or more writers around a pipeline of annotator components. While the V3NLP Framework is built upon the APACHE UIMA [16] platform, it utilizes UIMA-FIT [17] and Leo [18] functionalities that simplify and work around the complexities inherent in UIMA.

Related Work

The task of identifying mentions of sexual trauma in electronic medical notes using NLP is related to work developing social phenotyping using VHA electronic medical records described by Gundlapalli et al. [19]. The dictionary used for the sexual trauma task has been curated by subject matter experts and VHA clinical providers, and constrained to only include mentions of direct evidence of sexual trauma.

Methods

The Patient Cohort

This work was conducted using the extraction of a cohort of 10,000 male and 10,000 female patients randomly chosen from veterans who served during recent wars in Afghanistan and Iraq who had also received medical services in VHA medical facilities for the first time during FY2010-FY2014. All outpatient electronic medical notes in the 12 months following the date of the MST screen were extracted.

The clinical notes were extracted from VHA's Corporate Data Warehouse using the Veterans Informatics and Computing Infrastructure (VINCI) [20].

Natural Language Processing Pipeline

This sexual trauma extraction task is accomplished by a dictionary lookup mechanism within a larger NLP pipeline. A UIMA pipeline [16] was assembled using components of the open source V3NLP framework [15]. UIMA pipelines are composed of a series of annotators, where the output of one annotator is in turn the input to the next. The annotators chosen at the front of the sexual trauma pipeline decompose the text into constituent document element parts [21] including sections [22], content headings, lists, sentences, lines, tokens, slots and their values, questions and their answers, and check boxes, as well as

other boiler-plate entities. The dictionary lookup annotator provides single and multi-word term annotations (more about the dictionaries employed in a subsequent section). An annotator created for the sexual trauma task was employed to transform term annotations with sexual trauma categories into sexual trauma annotations. An annotator using rules from conTEXT [23] provided assertion attributes (negation, assertion, subject, hypothetical, conditional, historical) to sexual trauma annotation, followed by a filtering annotator to remove intermediate, but no longer needed, annotations. Figure 1 shows the annotators that make up this pipeline.

Term Identification Annotator

The term identification annotator creates term annotations from longest matching spans within sentence boundaries. This is the dictionary lookup portion of the pipeline. Term lookup uses the same lookup algorithm described in Divita et al [24]. Dictionary terms include one or more categorizations or semantic types. For this task, multiple dictionaries are used, a distinctive V3NLP Framework functionality. General and medical terminology is covered through the use of the SPECIALIST Lexicon [25]. The identification of general and medical terminology is used to absorb multi-word terms such as "MST Survey" that would otherwise cause ambiguity and fallacious matches if seen as individual words.

A dictionary of psychosocial risk factors created by domain experts in the field of homelessness formed the initial lexicon. Sexual trauma and related terms were extracted and further curated by subject matter experts with an understanding of sexual victimization. Additional terminology was found from traversing real-world usage from n-grams from a representative corpus [26]. In all, the domain dictionary includes 96 terms.

Thirty-five negative evidence terms were employed to reduce false positives. These included items like *book* and *pamphlet*. Such mentions often indicated that the contents of a book or pamphlet was the subject of a discussion during individual or group therapy sessions for sexual trauma survivors rather than a reference related directly related to sexual trauma.

Addressing Boilerplated Content

VHA electronic medical notes are replete with boiler-plated text [27, 28], similar to other electronic medical record systems. Such text is telegraphic, underspecified shorthand used to convey meaning by shortening the lengthy narrative that would otherwise be required. There is a large amount of variety and variability to the boiler-plated content within clinical text as providers are given the freedom to develop their own templates to facilitate documentation. Such content includes check boxes, slots and their values, questions and their answers, and pre-written prose text that has been copied-and-pasted into the record. The assertion semantics of sexual trauma mentions found within check boxes, slots and their values, and questions is different for each of these entities. For instance, a MST mention contained within the content heading of a check box is only asserted if the box is checked. A MST mention found within the content heading of a slot and filler structure is only asserted if the value or filler is filled out and has a non-negative kind of value. The assertion semantics are similar for sexual trauma mentions within questions. Integrations of prior work in this area was extended to this pipeline [21].

Exposure to boiler-plates involving sexual trauma mentions pointed to a challenging issue of boiler-plate forms involving assertions that a topic was discussed, rather than a positive assertion that the patient had the condition. An early review from a small seeded set of notes identified a number of boiler-plate MST

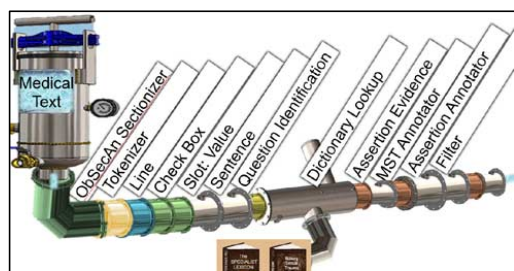


Figure 1- Sexual Trauma Pipeline

surveys. Mentions from these surveys were false positives more often than not. The domain lexicon was subsequently augmented with eight identifying patterns from these boilerplates, and labeled with an *ignore* category to filter mentions coming from these boilerplates out.

Sexual Trauma Annotator

The sexual trauma annotator, the only annotator custom-developed for this project, looks for sexual trauma terms as anchors within the bounds of each sentence or slot value structure. Once found, negative evidence was looked for within this same window. Anchors found without any negative evidence were made into sexual trauma annotations.

Scaling Up the Pipeline

The availability of a large set of records presented the opportunity to scale up the pipeline to efficiently handle hundreds of thousands of records in a reasonable amount of time. An application using the V3NLP Framework thread-based scale-up functionality [29] was used to wrap around and replicate the pipeline. This functionality allowed the process to start up with an initial number of pipelines, and, over time, add additional pipeline instances if there are enough resources available on the machine.

This task provided a use case to build out and include a mechanism to sample output for quality and efficacy purposes while the process is running. The sampling mechanism involves creating snippets around mentions at a specified sample rate. Visually, a snippet involves the line in the text a mention is found, plus lines above and lines below. Each snippet includes enough metadata to trace back to the location in the original document; it also aggregates information about the categorization of the mention, allowing analysis using just the snippets. Snippets are aggregated and put into VTT [30] formatted files to allow for quick human review and validation (Figure 2).

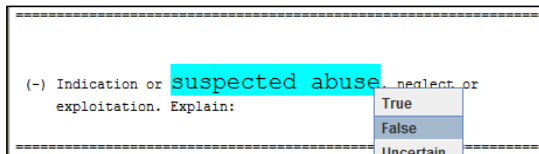


Figure 2- VTT Formatted Snippet Around a Boilerplate Question.

Evaluation

Sexual trauma mentions have been noted to be a rare event in clinical records, precluding a de-novo human review of random records to find mentions, and also precluding an evaluation that identifies the mentions this algorithm would miss. Rather, the snippets around each asserted mention were created from this cohort to evaluate those that the NLP extracted. All asserted snippets were reviewed and classified by three reviewers. Each reviewed snippet was classified as either a true positive or false positive. This review gives a positive predictive value (PPV) for the method. PPVs were also calculated for the snippets. The snippet decisions were aggregated to the patient level to give patient level PPV metrics. Gender was carried along, allowing PPV metrics to be detailed by gender.

The reviewers indicated for each mention whether the mention indicated childhood sexual trauma, adult (non-military) sexual trauma, or military sexual trauma. These curated annotations now form a reference corpus to guide work to more specifically identify kinds of sexual trauma. The reviewers went through three iterative rounds of training on sets of 200 snippets and

guideline tuning until the inter-annotator agreement amongst them reached 0.80.

Results

Of the initial sample of 20,000 patients, 13,501 had electronic medical notes in the timeframe designated for this study. These patients had a total of 362,546 notes that were processed using the NLP pipeline. A total of 8,299 asserted mentions were extracted by the pipeline from 5,357 documents (1.5% of all documents) and 1,220 unique patients (9% of all patients who had electronic medical notes).

Performance Metrics of V3NLP Pipeline for Sexual Trauma Mentions

As shown in Table 1, the overall precision (or positive predictive value) for identifying positively asserted sexual trauma mentions as validated by a human review is 0.90. The performance was nearly double for identifying mentions from the free text of electronic medical notes of female patients as compared to male patients. When the mentions were aggregated to the patient level, the overall precision decreased to 0.71. The metric for female patients was superior as compared to male patients.

Table 1— Performance metrics of V3NLP pipeline in identifying positively asserted sexual trauma mentions

	True Positive (N)	False Positive (N)	Precision (PPV)
Mentions from free text			
Male	441	558	0.41
Female	6,943	357	0.95
Total	7,384	915	0.90
Patient level metrics			
Male	114	184	0.38
Female	751	171	0.82
Total	865	355	0.71

Observations

Mentions were found in 411 document types. Figure 3 shows the top 20 document types covering 50% of the mentions extracted by the V3NLP pipeline.

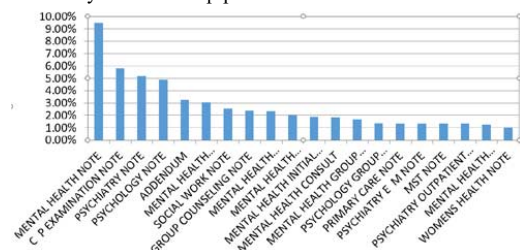


Figure 3 – Sexual Trauma Mentions By Standard Title

Performance Metrics

The scaled-out NLP pipeline processed 362,359 documents in 5.7 hours, processing at a rate of 57 ms per record. Ten concurrent pipelines were employed to achieve this performance.

Failure Analysis

Half the false positive snippets were reviewed and classified into eight categories (Table 2). Of note, failures having to do with negation account for 39% of the total failures. The majority of these were intertwined with boiler-plated (and in particular, check box and question formatted) text where the negation was too distant to be picked up, on subsequent lines or in a multi-column format (not currently handled), or using patterns not in our negation pattern expressions, such as *neg*. An additional 24% of the false positives came from statements that were overly broad in meaning: where two relevant and non-relevant mentions were used together with no ability to determine which one should be asserted (example of second type of false positive in Table 2). A prominent error was observed within the set of false positives: *MST* as an abbreviation for Mountain Standard Time in the US, an oversight to be accounted for going forward.

Table 2 – False Positive Analysis

Type of False Positive	%	Examples
Negation	39	<i>RAPE: In military {X} No e. SEXUAL ASSAULT Doesn't apply to Veteran {-} sexual harassment/abuse</i>
Overly broad	24	<i>Have you suffered from physical or sexual abuse? Yes Active Combat/POW/MST/War Zone: ... later to be displayed for Sexual Trauma</i>
Not referring to a person	13	<i>Awareness Month</i>
Not referring to the patient	8	<i>Veteran reports that she witnessed another person being sexually assaulted</i>
Headings	5	<i>PHYSICAL/SEXUAL ABUSE:</i>
Template issues	4	<i>2. sexual abuse {} yes {} no {}x} maybe</i>
Wrong Sense	3	<i>10/30/14 @ 1600 MST</i>

Discussion

A history of sexual trauma is not well documented in the electronic medical records of patients. However, it is important to identify sexual trauma survivors so that they may be offered appropriate trauma-informed services to mitigate adverse health-related outcomes.

This study demonstrates the feasibility of identifying and extracting positively asserted mentions of sexual trauma from the free text of electronic medical notes. The overall PPV of the NLP pipeline is high at the patient level (0.71) and even higher at the mention level (0.90), particularly when identifying sexual trauma in female patients.

With regard to which type of notes and providers are documenting sexual trauma in the free text of electronic medical notes in VHA, as expected, notes from mental health, psychiatry, and psychology providers were among those with the highest hit rate. A particular type of note which is specific to VHA is Compensation and Pension (C P examination note) that documents adverse experiences during military service that are then evaluated for disability compensation. In VHA, patients with a history of military sexual trauma are eligible for free care

and for monetary compensation; thus, it is expected that sexual trauma mentions would be prevalent in C P examination notes.

More than 60% of the false positives can be addressed. Though much work had gone into augmenting the negation patterns within the negation tool, the failure analysis shows that more effort is needed to address negation in particular. The majority of the negation failures will be addressed in the next iteration. The majority of the overly broad failures included a narrow set of co-occurring concepts such as physical abuse. Business rules will be applied using the set of co-occurring terms with sexual trauma to address these challenges.

Follow-on Work

The negative snippets need to be reviewed to see what false negatives are being inadvertently missed by the NLP pipeline. These snippets have been generated and human review is planned. When completed, this body of work will provide the true negatives and false negatives counts needed to compute precision/recall metrics.

Classifications of the sexual trauma annotations into adult, childhood, or military sexual trauma types will be added based on evidence statements found within the bounds of the snippet window. A preliminary analysis of such evidence in the manually classified snippets suggests that a high proportion of annotations may be successfully classified. Sexual trauma indicators from structured data including MST screening status will be correlated with these findings to provide a control vs positive cohort to calculate precision/recall/f-metrics.

Conclusions

Sexual trauma mentions are rarely noted in electronic medical notes of VHA patients. In our study, only 9% of patients positively asserted sexual trauma mentions in the free text of electronic medical notes and these were extracted from 1.5% of the document corpus. This paper demonstrates a technique to identify positively asserted sexual trauma mentions (words, phrases, and statements) from the free text of electronic medical notes of VHA patients using natural language processing. Documents with the highest hit rate of mentions as well as reasons for false positivity of the outputs are reported. The methods described here are generalizable and portable to documents from other health care systems outside VHA that use electronic medical record systems. This paper lays the foundation for techniques to further classify patients with sexual trauma with the intent to identify Veterans with military sexual trauma to better serve this vulnerable population.

Acknowledgements

This work is funded by grants VINCI HIR-08-204, CHIR HIR 08-374, ProWATCH grants HIR-10-001, HIR 10-002, and IIR 12-084 from the HS R&E. VA Informatics and Computing Infrastructure (VINCI) resources were used for this project. We acknowledge resources of the VA Salt Lake City IDEAS Center 2.0 (CIR 13-414) were used.

The views expressed in this paper are those of the authors and do not necessarily represent the views of the United States Department of Veterans Affairs or the United States government.

References

1. Kimerling, R., et al., *The Veterans Health Administration and military sexual trauma*. American Journal of Public Health, 2007. 97(12): p. 2160-2166.

2. Suris, A. and L. Lind, *Military sexual trauma a review of prevalence and associated health consequences in veterans*. Trauma, Violence, & Abuse, 2008. **9**(4): p. 250-269.
3. Sadler, A.G., et al., *Health-related consequences of physical and sexual violence: women in the military*. Obstetrics & Gynecology, 2000. **96**(3): p. 473-480.
4. Ullman, S.E. and L.R. Brecklin, *Sexual assault history and health-related outcomes in a national sample of women*. Psychology of Women Quarterly, 2003. **27**(1): p. 46-57.
5. Frayne, S.M., et al., *Medical profile of women Veterans Administration outpatients who report a history of sexual assault occurring while in the military*. Journal of women's health & gender-based medicine, 1999. **8**(6): p. 835-845.
6. Kimerling, R., et al., *Military sexual trauma and suicide mortality*. American journal of preventive medicine, 2016. **50**(6): p. 684-691.
7. Brignone, E., et al., *Differential Risk for Homelessness Among US Male and Female Veterans With a Positive Screen for Military Sexual Trauma*. JAMA psychiatry, 2016. **73**(6): p. 582-589.
8. Luterek, J.A., J.N. Bittinger, and T.L. Simpson, *Posttraumatic sequelae associated with military sexual trauma in female veterans enrolled in VA outpatient mental health clinics*. Journal of Trauma & Dissociation, 2011. **12**(3): p. 261-274.
9. Suris, A., et al., *Mental health, quality of life, and health functioning in women veterans differential outcomes associated with military and civilian sexual assault*. Journal of interpersonal violence, 2007. **22**(2): p. 179-197.
10. National Sexual Violence Resource Center, *Assessing Patients for Sexual Violence: A guide for Health Care Providers*. 2011.
11. Clark, C., et al., *Treating the trauma survivor : an essential guide to trauma-informed care*. 2015, New York, NY: Routledge, Taylor & Francis Group. xii, 166 pages.
12. Kimerling, R., et al., *The Veterans Health Administration and military sexual trauma*. Am J Public Health, 2007. **97**(12): p. 2160-6.
13. Friedman, L.S., et al., *Inquiry about victimization experiences: a survey of patient preferences and physician practices*. Archives of internal medicine, 1992. **152**(6): p. 1186-1190.
14. Rennison, C.M., *Rape and sexual assault: Reporting to police and medical attention, 1992-2000*. 2002: US Department of Justice, Office of Justice Programs Washington, DC.
15. Divita, G., et al., *v3NLP Framework: Tools to Build Applications for Extracting Concepts from Clinical Text*. eGEMs, 2016. **4**(3).
16. Ferrucci, D. and A. Lally, *UIMA: an architectural approach to unstructured information processing in the corporate research environment*. Natural Language Engineering, 2004. **10**(3-4): p. 327-348.
17. Ogren, P.V. and S.J. Bethard, *Building test suites for UIMA components*. in *Proceedings of the Workshop on Software Engineering, Testing, and Quality Assurance for Natural Language Processing*. 2009. Association for Computational Linguistics.
18. Patterson, O.V., T. Ginter, and S.L. DuVall, *Large scale clinical text processing and process optimization*.
19. Gundlapalli, A.V., et al., *Validating a strategy for psychosocial phenotyping using a large corpus of clinical text*. J Am Med Inform Assoc, 2013. **20** (e2): p. e355-64.
20. US Department of Veterans Affairs. *VA Informatics and Computing Infrastructure (VINCI)*. 2016 [cited 2016; Available from: <http://www.hsrd.research.va.gov/researchers/vinci/>].
21. Divita, G., et al. *Recognizing Questions and Answers in EMR Templates Using Natural Language Processing*. in *ICIMTH*. 2014.
22. Tran, L.-T.T., et al. *Scaling Out and Evaluation of OBSecAn, an Automated Section Annotator for Semi-Structured Clinical Documents, on a Large VA Clinical Corpus*. in *AMIA Annual Symposium Proceedings*. 2015. American Medical Informatics Association.
23. Chapman, W.W., D. Chu, and J.N. Dowling, *ConText: An algorithm for identifying contextual features from clinical text*. in *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. 2007. Association for Computational Linguistics.
24. Divita, G., et al., *Sophia: A Expedient UMLS Concept Extraction Annotator*. AMIA Annu Symp Proc, 2014. **2014**: p. 467-76.
25. Browne, A. *The SPECIALIST Lexicon and NLP Tools*.
26. Divita, G., et al., *Nora: A Vocabulary Discovery Tool for Concept Extraction*. Stud Health Technol Inform, 2015. **213**: p. 179-82.
27. Divita, G., et al., *PlateRunner: A Search Engine to Identify EMR Boilerplates*. Stud Health Technol Inform, 2016. **226**: p. 33-6.
28. Divita, G., et al., *Recognizing Questions and Answers in EMR Templates Using Natural Language Processing*. Stud Health Technol Inform, 2014. **202**: p. 149-52.
29. Divita, G., et al., *Scaling-out NLP Pipelines to Process Large Corpora of Clinical Notes*. Methods of Information in Medicine, 2015([Under review]).
30. Lu, C.J., G. Divita, and A.C. Browne. *Development of Visual Tagging Tool*. in *AMIA 2010 Annual Symposium; 11/2010*. 2010.

Address for correspondence

Correspondence: Guy Divita, MS, guy.divita@hsc.utah.edu or Adi V. Gundlapalli, MD, PhD, Chief Health Informatics Officer and Staff Physician, VA Salt Lake City Health Care System, Salt Lake City, Utah, USA adi.gundlapalli@hsc.utah.edu

General Symptom Extraction from VA Electronic Medical Notes

Guy Divita^{a,b}, Gang Luo, PhD^c, Le-Thuy T. Tran, PhD^{a,b}, T. Elizabeth Workman, PhD^{a,b},
Adi V. Gundlapalli, MD, PhD^{a,b}, Matthew H. Samore, MD^{a,b}

^a VA Salt Lake City Health Care System, Salt Lake City, Utah, USA

^b University of Utah School of Medicine, Salt Lake City, Utah, USA

^c Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, USA

Abstract

There is need for cataloging signs and symptoms, but not all are documented in structured data. The text from clinical records are an additional source of signs and symptoms. We describe a Natural Language Processing (NLP) technique to identify symptoms from text. Using a human-annotated reference corpus from VA electronic medical notes we trained and tested an NLP pipeline to identify and categorize symptoms. The technique includes a model created from an automatic machine learning model selection tool. Tested on a hold-out set, its precision at the mention level was 0.80, recall 0.74 and an overall f-score of 0.80. The tool was scaled-up to process a large corpus of 964,105 patient records.

Keywords:

Natural Language Processing; Machine Learning; Diagnosis

Introduction

There is need for more granular data, in particular, signs and symptoms, from the medical record in clinical and research domains such as quality improvement, population health metrics, patient recruitment for clinical trials, surveillance for adverse events, post-marketing surveillance, genomic medicine (genotype-phenotype associations), and epidemiologic studies. Structured data elements holding signs and symptoms have been shown to be underreported in clinical data repositories[1]. Advances in natural language processing (NLP) have begun to unlock information from the free text of medical notes[2]. This paper reports on an effort to extract signs and symptoms from clinical text using NLP. While prior work focused on specific signs and symptoms or worked within limited domains, this effort is more general.

Symptom Definition

Symptoms are the essence of the patient's experience of illness. A medical encounter starts with what a patient conveys to his or her provider in the form of symptoms and concerns. The traditional distinction between symptoms and signs is that symptoms are subjective experiences whereas signs are objectively observed. A more formal definition of a symptom is that it is "a bodily feature of a patient that is observed by the patient and is hypothesized by the patient to be a realization of a disease [3]. An operational criterion of a symptom includes the elements of patient experience, abnormal characteristics, and clinical relevance.

Symptoms are captured in the medical record as reported by providers. The patient's own words are filtered through the experience of the medical provider. Moreover, not all of the symptoms expressed by the patient are entered into the medical record [4]. Often, there is paraphrasing and summarization. When a symptom is difficult to paraphrase or capture in a single concept, it may be quoted verbatim in the record using quotation marks. Once the medical documentation is complete and the encounter is closed, there are only inferred references to the symptoms. Some, but not all symptoms are reflected in problem lists or ICD-9-CM coding of the medical encounter[5] or inferred on the basis of prescribed medications. Forbush noted that while problem lists included on average three symptoms per document, six symptoms on average are mentioned in the clinical note. Most documentation of symptoms is in the form of free text in clinical notes.

Common to other NLP extraction tasks, this work recognizes symptoms by a dictionary lookup methodology but with a large (over 92k) symptom concept dictionary. Additional notable methods described in this paper include recognizing common lexical patterns indicative of symptom phrases, recognizing only asserted mentions and an optimized machine learned component to filter out fallacious symptom phrases.

The impetus for this work revolves around improving patient care for veterans. US military personnel who have served in combat theaters experience various symptoms and illnesses attributable to their deployment [6, 7]. Of the most common conditions noted in administrative data of recently returned combat veterans are "non-specific signs and symptoms" represented by ICD-9-CM codes 780-799 [7]. While this is an appropriate starting point for epidemiologic studies, there is a need to identify symptoms in free text to address the true extent of post-deployment illnesses among Veterans seen in US Department of Veterans Affairs (VA) medical facilities.

More broadly, tracking and assessing the presence of symptoms is useful for surveillance of syndromes[8], staging of disease, and evaluation of treatment response. Phenotyping, which involves the characterization of a set of clinical features, is incomplete without the inclusion of the patient's subjective experience.

The objective of this project was to develop a natural language processing (NLP) pipeline that reliably identifies and extracts mentions of any positively asserted symptoms from the free text of clinical notes. We also address challenges facing current information extraction techniques such as the vast heterogeneity of expression and boilerplating commonly seen in electronic medical records.

Related Work

Extracting concepts from the free text of medical records (clinical text) has been the holy grail of NLP researchers. The challenges of processing clinical text over biomedical text have resulted in slow progress over the years[9]. Starting with outpatient and emergency department encounters, efforts have been underway to process the free text associated with chief complaint data, problem lists for continuity between visits, family history, chest x-ray reports, pathology reports and discharge summaries[10]. Several studies have focused on the free text of the medical encounter (both outpatient and inpatient) in looking for clues to adverse events or for bio surveillance[11]. Limited studies have focused on signs and symptoms associated with specific diseases or conditions; these include infectious diseases such as pneumonia[12] and influenza[13], and cancer staging[14]. Major impetuses to advance NLP of clinical text have been the serial i2b2 challenges. The 2010 challenge focused on problems, assertions and relationships[15]. Dligach, et al.[16] mentioned symptom extraction as a component of discovering body site and severity within clinical texts via cTAKES, but the signs and symptoms were not the focus of this work. More recently, Roberts[17] describes identifying symptom mentions through semantic categorization, extracting patterns that involve spatial relations between disorders and anatomical structures from well-formed prose. However, a very limited number of studies that have focused on symptoms expressed by patients in the body of the electronic note; there are virtually no studies on looking broadly at symptoms across sets of patients.

Methods

Reference Document Corpus

A sample of 948 records were extracted from a cohort of 6 million patient records from Veterans that had recently returned from deployment in Afghanistan and Iraq. The clinical notes were pulled from the VA's Corporate Data Warehouse (CDW) using the Veterans Informatics and Computing Infrastructure (VINCI) [18]. The records extracted were from 164 pre-selected document types. These records were human annotated to identify 5,819 positively asserted symptoms. Forbush describes the corpus characteristics and the annotation task.[5]. This corpus was divided into a training set and a hold-out testing set.

Natural Language Processing Pipeline

V3NLP Framework described in Divita et al[19] was used to build an NLP pipeline. V3NLP Framework is a framework built upon the Apache UIMA project[20].

This symptom extraction task is accomplished by a symptom dictionary lookup mechanism augmented with a statistical machine-learning filter. A UIMA pipeline was assembled using V3NLP framework components. UIMA pipelines are composed of a series of annotators, where the output of one annotator is in turn the input to the next. The annotators chosen at the front of the symptom pipeline decompose the text into constituent document element parts[21] including sections, content headings, lists, sentences, phrases, lines, tokens, slots and their values, questions and their answers, and check boxes, as well as other boilerplate entities. Additional annotators are included to add relevant features that will enable the downstream machine learning annotators to make an informed decision about whether a potential symptom is a true symptom or not. These annotators include a part of speech tagger and multi-word term identification to identify symptoms and non-symptoms. An an-

notator was created specifically for this task to identify potential symptoms by rules and patterns formed from annotations created by the dictionary lookup and document decomposition. The ConText assertion (negation, assertion, subject, hypothetical, conditional, historical) annotator [22] was included to add assertion attributes to potential symptoms to filter out negated and hypothetical symptom mentions such as *denies pain*, and *prn dizziness*.

A tail-end annotator was created for this task that employs a machine-learned model trained on 65 features gleaned from the upstream annotators. Figure 1 shows the production pipeline. The subsections that follow here describe the novel annotators within the symptom pipeline.

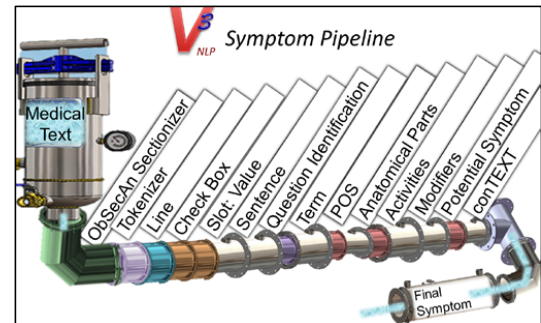


Figure 1 - V3 NLP Symptom Pipeline Used for Information Extraction from Free Text of VA Electronic Medical Notes

Document Element Decomposition Annotators

Section Identification

Section identification is accomplished by a wrapper around OBSecAn[23], which is a sectionizer built from the attributes of what makes up sections from a database of 35,000 document templates used within the VA's Veterans Health Information Systems and Technology Architecture (Vista) system.

Term identification Annotator

The term identification annotator creates term annotations from longest matching spans within sentence boundaries. This is the dictionary lookup portion of the pipeline. Term lookup uses the lookup algorithm described in Sophia, an Expedient Concept Extraction Tool[21]. Tokens are looked up from right to left across a sliding window, matching longest matching chains of tokens from an index that is similarly composed of reverse order tokens from terms.

Terms within the dictionaries include one or more categorizations or semantic types. For this task, multiple dictionaries are used, a distinctive v3NLP functionality. General and medical terminology is covered through the use of the SPECIALIST Lexicon. The identification of general and medical terminology is used to absorb multi-word terms such as "pain scale", that would otherwise cause ambiguity and fallacious symptoms if seen as individual words. A dictionary of 92,000 concepts (122,941 symptom forms) was created from Unified Medical Language System (UMLS) sources for this task, described by Tran[24]. Terms within this resource are tagged with a symptom category along with a set of 15 organ system sub-categories. A dictionary of idiosyncratic symptom phrases and symptoms not covered by the symptom dictionary (but seen in training data) is also employed. Terms from this resource are tagged with just the symptom category. A dictionary of symptom exceptions, or pertinent negatives, is included as a convenient way

to quickly incorporate exceptions for specific purposes. Such was needed to address failures seen from the training set.

Potential Symptom Annotator

The potential symptom annotator is a rule-based method that identifies those terms in the text marked with a symptom category and creates a potential symptom if these instances are not observed to be in content headings. The method looks for symptom mentions identified in the dictionary lookup within a window of a sentence and no evidence that would indicate the mention is not a symptom before promoting the symptom mention to a potential symptom. There are nearly 123 thousand strings that could make up possible symptom mentions. Thus far, this technique wildly over-generates potential symptoms.

Addressing Boilerplated Content

Clinical records are replete with boilerplated text. Such text is telegraphic, underspecified shorthand used to convey meaning by shortening the lengthy narrative that would otherwise be required. There is a large amount of variety and variability to the boilerplated content within clinical text. Such content includes check boxes, slots and their values, questions and their answers, and pre-written prose text that has been copied-and-pasted into the record. The assertion semantics of symptom mentions found within check-boxes, slots and their values and questions is different for each of these entities. For instance, a symptom contained within the content heading of a check-box is only asserted if the box is checked. A symptom mention found within the content heading of a slot and filler structure is only asserted if the value or filler is filled out and has a non-negative kind of value. The assertion semantics are similar for symptom mentions within questions. Integrations of prior work in this area was extended to this pipeline[21].

Additional Annotators

Previous work in this area has described colorful and often vague descriptions of symptoms [5]. Such descriptions include mentions of an anatomical location. An additional symptom pattern includes a normal activity and modifiers to that normal activity with some negative or pathologic connotation. Symptom patterns also often included some indication of severity and duration. A similar insight was further observed in reviewing false positives: a large portion of them involved an activity with some kind of positive modifier. For instance, mentions of *sleep* prefaced by *improved* were seen as false positives, but not poor sleep. Those observations led to the compilation of a dictionary of modifiers, activities, and anatomical locations from UMLS resources and the creation of annotators for each.

We used multiple methods to build these resources. To form a high-level list of anatomical locations, we extracted UMLS concept unique identifiers (CUIs) from the Consumer Health Vocabulary (CHV) project files[25] and mapped them to the corresponding terms in the 2014 SNOMED CT terminology. These terms were hand-curated to identify the surface type of terms one finds in a patient's symptom description. These were augmented with additional terms from the CHV's last terminology, found in the 2011AA UMLS Metathesaurus release. To locate normal activities that also intersect with findings and functions, we extracted terms from almost every English vocabulary in the 2014AA Metathesaurus UMLS release that had a personal behaviors semantic type (Activity, Behavior, Daily or Recreational Activity, Individual Behavior, or Social Behavior) and then terms for semantic types Finding and Organism Function. The final list consisted of words occurring in both of these groups. For the modifiers list, we extracted every adverbial term from the UMLS Specialist Lexicon LRAGR file (2014AA release). These were augmented with terms extracted from Patients Like Me symptom descriptions[26] and other

symptom descriptions on the Internet. Those activities, modifiers, and anatomical locations within the sentence that included a potential symptom were added as features to the machine learning.

Machine Learning Annotator: Training

Initially, the dictionary and rule based mechanisms produced approximately nine false symptom mentions for each true symptom mention. An additional mechanism using the surrounding context was needed to filter down the false positive mentions. An annotator was developed to create Weka ARFF data rows filled with the feature values needed to train Weka machine learning models.[27] This annotator was placed at the tail end of the training pipeline building an ARFF training row for each mention found, noting if the mention also overlapped a human marked symptom. The subsequent ARFF file involving 16,353 training rows (5,819 positive examples, 10,534 negative examples) was used to create a machine learned model based on features and whether a human annotation overlapped the mention.

We used the automatic machine learning model selection tool built by Luo et al.[28, 29] to systematically test every classification algorithm in Weka and tune hyper-parameters. Compared to other similar ones, this tool can greatly reduce search time and classification error rate[29].

A technical note here: all mentions from all 948 notes were used to create the initial ARFF file. The rows from the ARFF file were then randomized, separated into a training set consuming 90% of the examples and a hold-out 10% used for testing.

Machine Learning Features

Features were chosen on the basis of adding evidence to identify a possible symptom as a true symptom. Five words to the left and to the right of the potential symptom and their respective parts of speech, and the part of speech of the potential symptom are included. In earlier versions, these were all grouped into a bag-of-words vector. The current iteration includes a feature for each of the fifth, fourth, third, second and first words to the left and right. Each feature includes enumerated values for what words appear in that position keeping words that appear more than 2 times in that position. Positional features intended to capture boilerplate clues are included, such as if the symptom appears in a checkbox, slot, value, question, list, or sentence, and if it is within a section. Also included are features associated with section information including the section name, if the line that the symptom appears in has been indented, and if the line includes camel case or all upper case. Activity, modifier, and anatomical part features are included based on the afore mentioned insight. Also included is the assertion status of the symptom. An analysis of the attributes that contributed most to the outcome revealed that the symptom words, followed by symptom category, section name, the fourth, and fifth words to the left and second and third words to the right were the most salient attributes.

Production Pipeline

A final annotator was developed that mirrored the machine learning annotator in that it creates Weka instances around potential symptoms, which are subsequently passed through the Weka trained model to classify whether they are true or not. Symptom annotation instances are created for those that are classified as true symptoms.

Results

The automatic machine learning model selection tool[28, 29]selected support vector machine coupled with stochastic

gradient descent as the classification algorithm. Tested on five iterations of different randomly assigned 90% training, 10% hold-out mentions, this model held a consistent performance for identifying asserted symptoms. Table 1 shows the information retrieval metrics for this model.

The model was folded into the NLP tool, scaled-up and run on a larger set of 964,105 records randomly chosen from the larger OEF/OEF cohort. The process ran 32 concurrent pipelines and took 11 hours to run, at an average speed of 40 ms per record. In all, 59,412 symptom mentions were found from 19,914 documents from 10,397 patients. Figure 2 shows the distribution of organ system classes of the symptoms found in this cohort.

Table 1 - SGD Model Performance on Hold-out Set

Class	Precision	Recall	F-Measure
Symptom	0.80	0.74	0.80

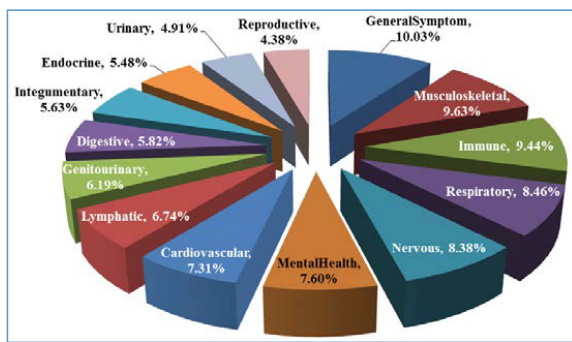


Figure 2- OEF/OIF Patient Symptom Distribution by Organ System

Failure Analysis

Snippets were generated around the false positives, and a sampling of 200 of these were categorized by type of failures. Issues in recognizing the numerous ways a mention can be negated are far more prevalent than any other type of error (57%). There were a number of false positives that could be plausibly true (16%). The rate of these are explainable due to the difficulty of the human annotation task. There were false positives caused by incorrectly parsing templates (5%) and checkboxes (5%). There was a smattering of other issues (21%) that could not easily be classified.

Discussion

Symptom extraction brings with it many challenges. Among them, a similarity and a continuum in context between signs, symptoms, findings, and diagnoses, making the distinction between these via explicit dictionary lookup and rules difficult. The term *depression* is a good example. It occurs 1160 times in this corpus. The depression mentions included references to patient reported symptoms, to provider observations and findings as well as the provider diagnoses. Our attempts at such without a machine learning component were disappointing. An additional requirement was the need to create a curated lexicon that significantly extended the pre-existing resources within UMLS. This additional resource was necessary to remedy the incompleteness of relevant terms within UMLS and to resolve the inconsistent distribution of symptoms across multiple UMLS semantic types. Moreover, our curated resource made it feasible to classify symptoms according to sub-type and organ system.

While the pipeline was developed specifically for VA medical notes, the general principles would be applicable to other large health care systems with commercial EMR's that contain free text and semi-structured notes with templates. The technique is useful for applications extracting patient described indications, useful for adding to the phenotype for conditions. The lessons learned with regards to document element decomposition and identification of slot value pairs would also be portable and generalization to other settings where EMRs are used.

Limitations

We were not able to accurately calculate metrics at the document or patient level due to randomizing the mentions before splitting the training and testing sets.

Despite related work that has greatly expanded the ConText patterns and sped up the application of the algorithm, negation continues to be the greatest source of false positives, at a rate of 60%.

Identifying a potential symptom is challenging for several reasons. The first is the observation that a large set of symptoms within the symptom dictionary are concepts that are a *finding* as well as a *normal* behavior, activity, or function. Such forms were observed to be a large portion of initial false positives. The heterogeneity of document types and the frequency and variety of boilerplated semi-structured elements continues to be troublesome. Despite the use of document element decomposition annotators, review of VA electronic medical notes reveals that symptom mentions are frequently found in telegraphic, boilerplated lists, check boxes and questions.

The practice of using terms denoting symptoms with both an activity and a modifier also poses a challenge for information extraction. No easy mechanism was identified to mark the modifier polarity; such information would be of benefit for future iterations of the pipeline.

The slot value annotator along with the question and answer annotator need more refined techniques to catch idiosyncratic formats, easy to understand visually, but difficult to generalize into patterns and rules. Beyond this, it should be noted that the training set contains inconsistencies with how some boilerplated sections were annotated or not annotated.

Word sense disambiguation, a challenge seen with the many acronyms and abbreviations, was not directly addressed in this pipeline. Co-reference resolution was partially addressed within the ConText algorithm which attributes if the symptom mention is attributed to the patient or other entity. Other than failures due to who to attribute the symptom to, co-reference resolution was not observed to be a point of failure. Neither issue rose to the levels of failure that negation or adequately parsing though check-boxes and questions currently pose.

The split for cross-validation was performed at the mention level. This may lead to some documents having mentions distributed into both the training set and the test set. Since a document may contain several occurrences of the same symptom, this is liable to result in an optimistic evaluation of the classifier results.

Future Work

This is being deployed in several applications where the focus is narrowed to specific conditions. We should learn how well the tool identifies specific kinds of symptoms from these studies. Recently added VINCI tools should allow us to compare our technique with cTAKES and CLAMP surrogates.

Conclusion

We have developed a technique to identify a variety of signs and symptoms within a wide range of document types. An exhaustive algorithm was used to find the most robust machine learning model to train with. The technique has been efficacy benchmarked with an f-metric of 0.80 against a hold-out set of symptom mentions. The technique was scaled-up and run across close to one million records. The pipeline and application is distributed under an Apache License at <http://inlp.bmi.utah.edu/redmine/docs/v3nlp-framework>

Acknowledgements

This work is funded by the United States Department of Veterans Affairs, OR&D, Health Services Research and Development grants VINCI HIR-08-204, CHIR HIR 08-374, ProWATCH grants HIR-10-001 and HIR 10-002. VA Informatics and Computing Infrastructure (VINCI) resources were used to develop and run this work.

The views expressed in this paper are those of the authors and do not necessarily represent the views of the United States Department of Veterans Affairs or the United States government.

References

- Pakhomov, S., et al., *Agreement between patient-reported symptoms and their documentation in the medical record*. The American journal of managed care, 2008. **14**(8): p. 530.
- Nadkarni, P.M., L. Ohno-Machado, and W.W. Chapman, *Natural language processing: an introduction*. J Am Med Inform Assoc, 2011. **18**(5): p. 544-51.
- Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an ontological treatment of disease and diagnosis*. Summit on translational bioinformatics, 2009. **2009**: p. 116.
- Pakhomov, S.V., et al., *Agreement between patient-reported symptoms and their documentation in the medical record*. Am J Manag Care, 2008. **14**(8): p. 530-9.
- Forbush, T.B., et al., "Sitting on pins and needles": characterization of symptom descriptions in clinical notes". AMIA Jt Summits Transl Sci Proc, 2013. **2013**: p. 67-71.
- Smith, T.C., et al., *Chronic multisymptom illness: a comparison of Iraq and Afghanistan deployers with veterans of the 1991 Gulf War*. Am J Epidemiol, 2014. **180**(12): p. 1176-87.
- Spelman, J.F., et al., *Post deployment care for returning combat veterans*. J Gen Intern Med, 2012. **27**(9): p. 1200-9.
- Elkin, P.L., et al., *Comparison of natural language processing biosurveillance methods for identifying influenza from encounter notes*. Annals of Internal Medicine, 2012. **156**(1_Part 1): p. 11-18.
- Meystre, S.M., et al., *Extracting information from textual documents in the electronic health record: a review of recent research*. Yearb Med Inform, 2008. **35**: p. 128-44.
- Demner-Fushman, D., W.W. Chapman, and C.J. McDonald, *What can natural language processing do for clinical decision support?* Journal of biomedical informatics, 2009. **42**(5): p. 760-772.
- Chapman, W.W., et al., *Natural language processing for biosurveillance*, in *Infectious Disease Informatics and Biosurveillance*. 2011, Springer. p. 279-310.
- Chapman, W.W., et al., *Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap*. Medinfo, 2004. **11**(Pt 1): p. 487-91.
- South, B.R., et al. *Optimizing A syndromic surveillance text classifier for influenza-like illness: Does document source matter?* in *AMIA Annual Symposium Proceedings*. 2008. American Medical Informatics Association.
- Nguyen, A.N., et al., *Symbolic rule-based classification of lung cancer stages from free-text pathology reports*. Journal of the American Medical Informatics Association, 2010. **17**(4): p. 440-445.
- Uzuner, Ö., et al., *2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text*. Journal of the American Medical Informatics Association, 2011. **18**(5): p. 552-556.
- Dligach, D., et al., *Discovering body site and severity modifiers in clinical texts*. Journal of the American Medical Informatics Association, 2013: p. amiajnl-2013-001766.
- Roberts, K., et al. *Automatic Extraction and Post-coordination of Spatial Relations in Consumer Language*. in *AMIA Annual Symposium Proceedings*. 2015. American Medical Informatics Association.
- US Department of Veterans Affairs. *VA Informatics and Computing Infrastructure (VINCI)*. 2015 [cited 2015; Available from: http://www.hsrd.research.va.gov/for_researchers/vinci/].
- Divita, G., et al., *v3NLP Framework: Tools to Build Applications for Extracting Concepts from Clinical Text*. eGEMs, 2016. **4**(3).
- Ferrucci, D. and A. Lally, *UIMA: an architectural approach to unstructured information processing in the corporate research environment*. Natural Language Engineering, 2004. **10**(3-4): p. 327-348.
- Divita, G., et al., *Recognizing Questions and Answers in EMR Templates Using Natural Language Processing*. Stud Health Technol Inform, 2014. **202**: p. 149-52.
- Chapman, W.W., D. Chu, and J.N. Dowling, *ConText: an algorithm for identifying contextual features from clinical text*, in *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*. 2007, Association for Computational Linguistics: Prague, Czech Republic. p. 81-88.
- Tran, L.-T.T., et al., *OBSeAnnot: An Automated Section Annotator for Semi-structured Clinical Documents*, in *Medinfo*. 2015.
- Tran, L.-T.T., et al., *Exploiting the UMLS Metathesaurus for Extracting and Categorizing Concepts Representing Signs and Symptoms to Anatomically Related Organ Systems*. Journal of Biomedical Informatics, 2015.
- Zeng, Q.T., et al., *Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system*. BMC Med Inform Decis Mak, 2006. **6**: p. 30.
- Patients Like Me*. 2014; Available from: www.patientslikeme.com.
- Frank, E., et al., *Data mining in bioinformatics using Weka*. Bioinformatics, 2004. **20**(15): p. 2479-81.
- Luo, G., *MLBCD: a machine learning tool for big clinical data*. Health information science and systems, 2015. **3**(1): p. 1.
- Zeng, X. and G. Luo, *Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection*.

Addresses for Correspondence

Guy Divita, MS, guy.divita@hsc.utah.edu or Adi V. Gundlapalli, MD, PhD, Chief Health Informatics Officer and Staff Physician, VA Salt Lake City Health Care System, Salt Lake City, Utah, USA adi.gundlapalli@hsc.utah.edu

An Integrated Surveillance System to Examine Testing, Services, and Outcomes for Sexually Transmitted Diseases

Brian E. Dixon^{a,b}, Guoyu Tao^c, Jane Wang^b, Wanzhu Tu^{b,d},
Sarah Hoover^b, Zuoyi Zhang^b, Teresa A. Batteiger^d, Janet N. Arno^d

^a Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA,

^b Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA

^c Division of STD Prevention, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

^d Indiana University School of Medicine, Indianapolis, Indiana, USA

Abstract

Despite laws that require reporting of sexually transmitted diseases (STDs) to governmental health agencies, integrated surveillance of STDs remains challenging. Data and information about testing are fragmented from information on treatment and outcomes. To overcome this fragmentation, data from multiple electronic systems spanning clinical and public health environments were integrated to create an STD surveillance registry. Electronic health records, disease case records, and birth registry records were linked and then stored in a de-identified, secure server for use by health officials and researchers. The registry contains nearly 6 million tests for 628,138 individuals over a 12-year period. The registry supports efforts to understand the epidemiology of STDs as well as health services and outcomes for those diagnosed with STDs. Specialized disease registries hold promise for collaboration across clinical and public health domains to improve surveillance efforts, reduce health disparities, and increase prevention efforts at the local level.

Keywords:

Sexually Transmitted Diseases; Registries; Public Health Informatics

Introduction

Sexually Transmitted Diseases (STDs)

Undiagnosed and untreated sexually transmitted disease (STD) is associated with adverse outcomes such as infertility, increased HIV transmission and acquisition, and adverse pregnancy outcomes. Several STD health services are recommended by the Centers for Disease Control and Prevention (CDC) to protect the reproductive and sexual health of young men and women. Recommendations include: annual chlamydia and gonorrhea screening of sexually active women ≤ 24 years, pregnant women, and older at-risk women; chlamydia and gonorrhea screening of anatomic sites of exposure (urethral, rectal, or pharyngeal) of men who have sex with men (MSM); retesting of all infected persons after treatment for chlamydia or gonorrhea; and syphilis testing of pregnant women as well as sexually active MSM [1].

Surveillance of STDs and STD Services

A core function of public health is the assessment of disease prevalence and burden as well as the utilization of health care services, also referred to as public health surveillance [2].

Ministries around the globe seek to perform surveillance on STDs as well as the utilization of STD health services. They further seek to monitor the quality of health services received by at-risk groups, assess adherence to recommendations for chlamydia and gonorrhea testing and retesting, syphilis testing, test results, patient and partner treatment, and the incidence of adverse outcomes related to STDs.

Assessing STD prevalence, burden and utilization of health services is challenging, because available data sources are limited by small sample sizes, incomplete demographic information, cross-sectional design, insufficient periods of follow-up time, and incomplete information about the services provided [3]. Access to a longitudinal data source with complete demographic and clinical information for individual patients is challenging for public health agencies and researchers, especially in the United States, given the fragmented delivery of care in public and private settings. Furthermore, there are even fewer data sources that capture an entire geographic community as opposed to a population defined by a single institution that provides care or insurance (such as a managed care population). While data sources such as population health surveys provide partial information, none have been able to provide all the information required to assess community access, utilization and quality of services, and the incidence of adverse outcomes following an STD.

Specialized Disease Registries

Centralized data registries have become important informatics tools for surveillance and research in a variety of public health contexts, including cancer treatment [4], immunization programmes [5], and injury prevention [6]. In fact, expanded health policies in the United States, referred to as “meaningful use” criteria for electronic health record (EHR) systems, include disease registries as a ‘public health’ criterion for the years 2013-2018 [7]. These policies encourage providers to submit patient-level information to specialized registries.

Once populated, disease registries can be reused for a variety of purposes, including clinical performance improvement, surveillance of disease incidence, and research on the utilization of health services [8; 9]. In essence, disease registries serve as integrated surveillance systems that support a wide range of clinical and public health functions.

Research Objective

Given the need for better community-level surveillance of STDs and STD health services as well as the past success of

other population disease registries, we sought to develop a longitudinal, comprehensive patient-centric registry to examine STDs and STD health services in a large metropolitan area. We hypothesized that the registry would support analysis of STDs and STD services as well as ongoing surveillance practice among public health agencies in that community.

Methods

We created a registry for all individuals tested for one of three STDs (chlamydia, gonorrhea, syphilis) between January 1, 2003, and December 31, 2014, by healthcare providers in the Indianapolis MSA (metropolitan statistical area). To create the registry, we gathered data from clinical and public health sources, linked individual patient records, and created a secure environment to facilitate collaborative access for surveillance and research. Our work occurred in partnership with local, state, and federal public health partners and was approved by the Institutional Review Board (IRB) at Indiana University.

Geography and Population Information

According to the 2010 census, Indiana ranked 15th among the states by population with just under 6.5 million residents. Consistent with national data, STDs are over-represented in racial and ethnic minorities (cases per 100,000 population). For example, the 2015 rate of gonorrhea among African Americans was 836 compared to the rate among Caucasian 87.7 and Hispanic individuals 85.0. The rates for chlamydia were 2234 for African Americans, 319 for Caucasians, and 545 for Hispanics for primary and secondary syphilis (26.8, 6.6 and 16.6, respectively).

The Indiana State Health Department (ISDH) STD Control Program divides the state's 92 counties into ten districts for morbidity reporting and disease intervention purposes. These district offices are the recipients of contracts with the STD Program for the state's approximately 30 disease intervention specialists. The Marion County Public Health Department (MCPHD) STD Control Program has responsibility for STD reporting in District 5, which includes Marion County (Indianapolis) and the seven surrounding counties: Boone, Hamilton, Hancock, Hendricks, Johnson, Morgan, and Shelby. This district makes up the majority of the Indianapolis MSA. District 5 (population of 1.7 million) and Marion County (population of 903,393) account for the largest share of Indiana's STD morbidity. In 2015, District 5 accounted for 39% of the state's chlamydia morbidity, 47% of the state's gonorrhea, and 60% of the state's primary and secondary syphilis. This partially reflects the district's racial health disparities, which is substantially more diverse than the state.

According to the CDC's 2015 STD Surveillance Report, Indiana reported a total of 28,886 cases of chlamydia and ranked 27th among states in rate (437.9/100,000), while Marion County ranked 25th among United States counties and independent cities at 949.3 cases/100,000 population. Indiana ranked 23th among states for gonorrhea with a case rate of 118.9/100,000 population, while Marion County ranked 16th among United States counties and independent cities in the rate of gonorrhea cases with 344.1 cases/100,000 population.

Residents of District 5 receive STD diagnostic and treatment services through the Bell Flower Clinic, the STD Control program of MCPHD, which also houses the District 5 reporting site. The program is operated by the Health and Hospital Corporation, which also operates MCPHD and safety net hospital for the county. The Bell Flower Clinic, therefore,

serves those at highest risk. Of the unique patients at the Bell Flower Clinic, 57% were African-American, 33% were Caucasian, and 7% were other. Seven percent were Hispanic, mostly of Mexican descent.

Data Sources

Data for the registry came from three distinct sources:

1. The Indiana Network for Patient Care (INPC), a regional health information exchange (HIE) network that contains longitudinal EHRs for patients who received care in the Indianapolis MSA.
2. MCPHD Bell Flower Clinic, the STD Control Program which houses an information system where disease investigators enter details about STD cases reported to public health for the Indianapolis MSA.
3. MCPHD Birth Registry, a vital records information system used at MCPHD to capture data on all births in Marion County, in which Indianapolis is located.

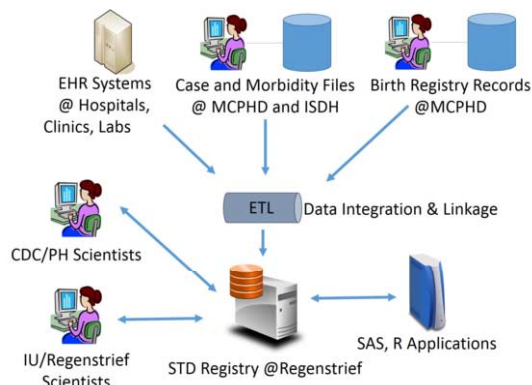


Figure 1 – Diagram depicting data sources, how data are integrated, and how data are accessed for surveillance

Indiana Network for Patient Care

The INPC is one of the largest community-based HIE networks in the United States [10; 11]. The INPC connects over 90 healthcare facilities, including hospitals, physicians' practices, pharmacy networks, long-term post-acute care facilities, laboratories, and radiology centers. The INPC maintains over five billion structured observations for over 12 million individuals; nearly one million electronic healthcare transactions are processed every day.

From the INPC, we extracted demographic data (e.g., gender, age, race, county of residence), STD laboratory testing data (e.g., lab test, date of test, result), co-morbidity data (e.g., pregnancy status, HIV status, ICD diagnoses) at time of STD test, encounter data (e.g., visit date, visit type), and medication history (e.g., drug name, drug class, date of dispense).

MCPHD Bell Flower Clinic and ISDH Morbidity Data

From the files at the Bell Flower Clinic, we extracted demographic data (e.g., gender, age, race), STD laboratory testing data (e.g., lab test, date of test, result), co-morbidity data (e.g., pregnancy status, HIV status) at time of the STD test, and medication information (e.g., drug name, drug class, date of dispense).

Because STD treatment may not be fully captured by the INPC, we extracted treatment of STD morbidity information

from the ISDH reporting database, SWIMSS (Statewide Investigating, Monitoring and Surveillance System).

MCPHD Birth Registry

The MCPHD Birth Registry contains records on all live births in hospitals and birthing centers in Marion County as reported by birth registrars. From the birth registry, we extracted pregnancy outcomes (e.g., date of delivery, infant weight, gestational age), STD laboratory testing data (e.g., lab test, date of test, result), and co-morbidity data (e.g., HIV status) at time of delivery.

Record Linkage, Integration, and Preparation

Data were integrated from the three distinct sources using a two-step process (Figure 1). First, individuals were linked across datasets. Next, data from each source was extracted and combined into a single, patient-centric data registry.

The INPC employs an advanced, probabilistic matching algorithm that matches patient identities using first name, last name, social security number (when available), date of birth, phone number (when available), and gender [12]. The two MCPHD datasets were independently linked to the INPC using the enterprise master person index (eMPI), based on that algorithm. Individuals in the MCPHD STD Case and Morbidity Files who did not match to an INPC individual were imported into the registry as new clients. Only data for individuals in the MCPHD Birth Registry who matched an INPC individual were imported from the vital records system.

Once patient identities were linked, longitudinal data for each unique individual were extracted, transformed, and loaded from the three sources into the registry. Each unique individual was given a de-identified or pseudonymised “client ID” that did not resemble his/her medical record number or any identifiers in the MCPHD datasets. Birth dates were transformed to ages and other identifiable information was removed. The ability to re-identify individuals exists to enable capture of new and updated information using a key between the medical record number and the client ID. Only the data manager at Regenstrief can perform data updates. Registry users cannot access such details to ensure confidentiality of records.

Data Management

The deidentified, linked registry datasets are hosted on a secure, virtual server at the Regenstrief Institute (Figure 1). The encrypted server is password-protected and managed by the technical services division at Indiana University (IU). The datasets are stored as a collection of interoperable data files, enabling them to be interpreted by all major analytical software tools. The data files require 20GB disk space.

Authorized users include public health scientists at the CDC and MCPHD as well as scientists working at Regenstrief and IU. The virtual environment affords users the opportunity to leverage a wide range of analytical tools, including SAS, R, and SPSS. Analyses of the data can be performed within the IU high-performance computing environment without necessitating download of the data onto local computers or drives. Output from the analyses, such as tables, charts, and graphs, can be downloaded from the servers to support in public health or academic reports.

Results

The registry contained 5,093,863 STD tests for 628,138 unique individuals collected over a 12-year period. In Table 1, we present the demographics of the individuals who were tested and those who tested positive for an STD in comparison to the overall demographics for the Indianapolis MSA. Although the area was well-balanced with respect to gender, a greater proportion of females were tested for STDs. This is likely due to clinical guidelines that recommend screening pregnant women and young, sexually active women for STDs. African American individuals were proportionately tested more and had a greater proportion of disease, than other races. These data highlight both a racial disparity in disease burden as well as the fact that providers are more regularly screening this population.

Table 1 – Demographics for individuals in the registry

Demo-graphic	Individuals Tested for an STD N=628,138	Individuals Positive for an STD N=119,751	Population of the MSA N=1,988,817
Gender			
Male	17.5%	25.3%	49.3%
Female	82.4%	74.6%	50.7%
Race			
African-American	25.0%	61.9%	15.3%
Caucasian	48.2%	25.1%	79.2%
Asian	0.4%	0.2%	2.9%
Hispanic	3.4%	2.7%	6.5%
Age			
5-17		10.4%	18.3%
18-24		40.0%	8.8%
25-44		25.8%	27.6%

In Figure 2, we summarize test results and positivity over time across all three STDs. Positivity is defined as the number of positive laboratory tests that confirm the presence of disease divided by the total number of lab tests analyzed. The overall volume of tests captured by the registry increased through 2011 then plateaued as result of the growth in the contributing data sources to the INPC from providers joining the HIE network to comply with ‘meaningful use’ incentives. Growth in the volume of data captured by the INPC resulted in decreased positivity; the number of negative tests grew by a factor of 3, while the number of positive tests increased by a factor of 2.5 (from 10,044 in 2003 to 25,606 in 2014). While the total number of positive STD cases grew dramatically, this growth is attributed to increased electronic lab reporting rather than an outbreak of disease.

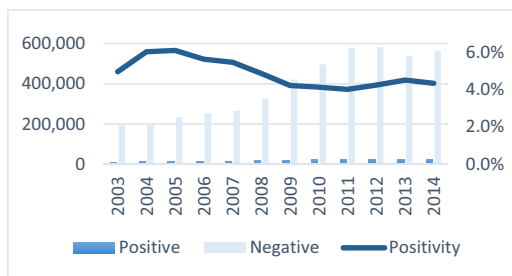


Figure 2 – Longitudinal test results and positivity for chlamydia, gonorrhea and syphilis. Test results are depicted as bars; positivity is depicted as a trend line

Discussion

By integrating three disparate sources of routinely collected clinical and public health data, we have created a novel registry containing longitudinal data on individuals tested for STDs in a large metropolitan area. The STD registry is an important public health informatics resource as it affords surveillance and research on STD testing, services, and outcomes. While each data source may exist independently in states and nations around the world, very rarely are EHR systems, vital records, and STD morbidity files linked and used to examine those tested or treated for STDs.

Most often, health departments only have access to positive laboratory results, which are required by law to be reported to health authorities [13]. While electronic laboratory reporting of positive test results improves the completeness and timeliness of public health reporting [14; 15], the lack of negative test reporting prevents health departments from examining whether individuals at risk for STDs are receiving recommended screening. Moreover, while health departments maintain both STD morbidity and vital record information systems, many health departments fail to link these data to examine outcomes for pregnant women or populations at risk for poor birth outcomes. Therefore, integrating clinical and public health datasets allows for an expanded evaluation of preventative services, clinical guidelines, and outcomes experienced by those with STDs.

Our work demonstrates the feasibility of creating a specialized STD registry for conducting surveillance and research. Building the registry further highlights three lessons for the biomedical and public health informatics communities. First, specialized registries that cross clinical and public health boundaries can be created in a way that preserves privacy and confidentiality. Second, record linkage is a crucial aspect of creating a registry. Third, health IT policies affect the breadth and depth of specialized registries.

Maintaining Privacy and Confidentiality of Health Data

Individuals and health organizations can be fearful of centralized, monolithic databases that contain protected health information [16]. Therefore, healthcare providers may be wary of releasing identifiable information to public health authorities, except when required by law to do so.

To create our registry, we leveraged the Regenstrief Institute, a neutral third party with experience in protecting health data. Regenstrief is a business associate with healthcare providers, public health authorities, and the INPC [11]. As a convening,

trusted partner, the Institute was able to bring clinical and public health organizations together to exchange identifiable data that could be linked and then de-identified for storage in a secure, common environment that affords surveillance and research by multiple users. The role of neutral third parties is supported by prior HIE research [17]; therefore, public health authorities should look to HIE networks or other third parties to support creating and maintaining specialized registries.

The Critical Role of Data Linkage

One of the most important and challenging aspects of creating the registry was record linkage. Linkage is important because uniquely identifying individuals is critical to pulling fragmented EHRs together for tracking an individual's STD testing and services longitudinally.

Because the original data sources independently maintain distinct, unique identifiers for individuals and the United States lacks a universal health identifier, there was no easy method to link individuals at the start of the project. While the probabilistic algorithm used by the INPC's eMPI is excellent, it is not flawless. Therefore, each public health source had to be independently matched to the INPC, and then the two matched sets had to be linked using a third round of matching. Due to the lack of universal identifiers, each round of matching involved some degree of manual review and a decision threshold for determining correct matches had to be established. The necessity of manual review prohibits automation and scaling of specialized registry creation.

One potential solution for others is a client registry (CR) [18]. A CR adjudicates identities across EHR and other health data systems, like vital records and morbidity information systems, producing a centralized MPI to link identities across data sources. CRs have been demonstrated in HIE networks emerging in several countries, including Rwanda [19]. The CR should be further studied and applied to specialized disease registries.

Robust Policies Facilitate Specialized Registries

The STD registry is but one example of a specialized health data registry. While a wide range of registries for injuries, vaccines, and diabetes existed before the HITECH (Health Information Technology for Economic and Clinical Health) Act of 2009, the "meaningful use" program's incentive for clinical providers to contribute data to a specialized registry encourages clinical-public health data exchange. Local health authorities struggle to receive data that are currently not covered under existing public health laws. While new laws can be written to require data exchange, health authorities have an opportunity to leverage existing policies, like HITECH, to work with clinical providers to create and sustain population health surveillance through specialized registries.

When creating registries, health authorities should consider the unique health needs of their jurisdiction. Community health assessments, an activity involving the gathering of input from a wide array of stakeholders on the important issues facing a community, are another opportunity to work with healthcare providers to identify key health issues that might benefit from a specialized disease registry. Diabetes may be a top priority in one nation, while hypertension might be a top priority in another jurisdiction. Working with healthcare providers to identify the health priorities of a community may lead to better participation in the registry as well as progress in "moving the needle" towards higher quality of care and outcomes for at-risk populations.

Future Directions for the STD Registry

The STD registry allows our team to explore many important questions relevant to public health practice and research. Our team is currently conducting the following analyses and plans to disseminate results in the coming year:

- **Utilization of STD Services:** Understanding where individuals present for STD services is critical for appropriately allocating available resources. Using the data available within the registry, we are examining testing locations and positivity rates of individuals to determine where individuals present for STD care and whether a positive result increases the likelihood of presenting to a specific location.
- **Testing and Outcomes for Pregnant Women:** Women should be screened and treated for STDs while pregnant. Using the available testing data for women who either tested positive for pregnancy or delivered a baby, we are examining the proportion who received an STD test; of those, which women were positive and the birth outcomes for women who tested positive.

In addition, we seek to expand the capacity for the registry to support other research and surveillance of STD testing, services, and outcomes. In the next year, we plan to link the registry to other unique public health datasets, including the Immigrant Tuberculosis and All Refugee Application (ITARA) database. This database includes information on Indiana state immigrant medical exams. Incorporating these data will facilitate an analysis of newly immigrated citizens for incidence as well as risk factors associated with STDs. The registry will continue to be hosted at Regenstrief for use by public health researchers as well as epidemiologists in local, state, and federal agencies.

Conclusion

Using multiple data sources, we successfully linked and integrated data relevant to the testing, treatment, and outcomes for individuals with STDs to create a specialized STD registry. Registries like this one are increasingly feasible to build using informatics approaches. Specialized disease registries are critical to understanding the epidemiology of disease and enable collaboration across clinical and public health domains to improve surveillance, reduce health disparities, improve health services for individuals with disease, and increase prevention efforts at the local level.

Acknowledgements

The research reported in this publication was supported by the Centers for Disease Control and Prevention (CDC), United States Department of Health and Human Services (HHS), under Contract Number 200-2011-42027. The content is solely the responsibility of the authors and does not necessarily represent the official views of CDC or HHS.

References

- [1] Centers for Disease Control and Prevention, 2015 Sexually Transmitted Diseases Treatment Guidelines, in, U.S. Department of Health and Human Services, Atlanta, GA, 2015.
- [2] L.M. Lee and S.B. Thacker, The cornerstone of public health practice: public health surveillance, 1961–2011, *MMWR. Surveillance*

summaries: Morbidity and mortality weekly report. *Surveillance summaries / CDC* **60 Suppl 4** (2011), 15–21.

- [3] M.B. Ivankovich, J.S. Leichter, and J.M. Douglas, Measurement of Sexual Health in the U.S.: An Inventory of Nationally Representative Surveys and Surveillance Systems, *Public Health Reports* **128** (2013), 62–72.
- [4] S. Mehra, R.M. Tuttle, M. Milas, L. Orloff, D. Bergman, V. Bernet, E. Brett, R. Cobin, G. Doherty, B.L. Judson, J. Klopper, S. Lee, M. Lupo, J. Machac, J.I. Mechanick, G. Randolph, D.S. Ross, R. Smallridge, D. Terris, R. Tufano, E. Alon, J. Clain, L. DosReis, S. Scherl, and M.L. Urken, Database and Registry Research in Thyroid Cancer: Striving for a New and Improved National Thyroid Cancer Database, *Thyroid* **25** (2014), 157–168.
- [5] Progress in immunization information systems - United States, 2012, *MMWR Morb Mortal Wkly Rep* **62** (2013), 1005–1008.
- [6] R.J. Mitchell, C.M. Cameron, and M.R. Bambach, Data linkage for injury surveillance and research in Australia: perils, pitfalls and potential, *Aust N Z J Public Health* **38** (2014), 275–280.
- [7] Medicare and Medicaid Programs; Electronic Health Record Incentive Program--Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. Final rules with comment period, *Fed Regist* **80** (2015), 62761–62955.
- [8] G.W. Hruby, J. McKiernan, S. Bakken, and C. Weng, A centralized research data repository enhances retrospective outcomes research capacity: a case report, *J Am Med Inform Assoc* **20** (2013), 563–567.
- [9] B.E. Dixon, E.C. Whipple, J.M. Lajiness, and M.D. Murray, Utilizing an integrated infrastructure for outcomes research: a systematic review, *Health Info Libr J* **33** (2016), 7–32.
- [10] P.G. Biondich and S.J. Grannis, The Indiana network for patient care: an integrated clinical information system informed by over thirty years of experience, *J Public Health Manag Pract Suppl* (2004), S81–86.
- [11] J.M. Overhage, The Indiana Health Information Exchange, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*, B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 267–279.
- [12] S.J. Grannis, J.M. Overhage, S. Hui, and C.J. McDonald, Analysis of a probabilistic record linkage technique without human review, *AMIA Annu Symp Proc* (2003), 259–263.
- [13] T.J. Doyle, M.K. Glynn, and S.L. Groseclose, Completeness of notifiable infectious disease reporting in the United States: an analytical literature review, *Am J Epidemiol* **155** (2002), 866–874.
- [14] J.M. Overhage, S. Grannis, and C.J. McDonald, A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions, *Am J Public Health* **98** (2008), 344–350.
- [15] B.E. Dixon, J.J. McGowan, and S.J. Grannis, Electronic laboratory data quality and the value of a health information exchange to support public health reporting processes, *AMIA Annu Symp Proc* **2011** (2011), 322–330.
- [16] R.V. Dhopeswarkar, L.M. Kern, H.C. O'Donnell, A.M. Edwards, and R. Kaushal, Health care consumers' preferences around health information exchange, *Ann Fam Med* **10** (2012), 428–434.
- [17] N.M. Lorenzi, Strategies for Creating Successful Local Health Information Infrastructure Initiatives in: U.S. Department of Health and Human Services, ed., Assistant Secretary for Policy and Evaluation, Washington, DC, 2003.
- [18] T.D. McFarlane, B.E. Dixon, and S.J. Grannis, Client Registries: Identifying and Linking Patients, in: *Health Information Exchange: Navigating and Managing a Network of Health Information Systems*, B.E. Dixon, ed., Academic Press, Waltham, MA, 2016, pp. 163–182.
- [19] Jembi Health Systems, Rwanda Health Enterprise Architecture (RHEA), in, 2012.

Address for correspondence

Brian E. Dixon, PhD
Regenstrief Institute, Inc., 1101 W. 10th St., RF 336
Indianapolis, IN 46202 USA
+1 (317) 278-3072
bedixon@regenstrief.org

Development and Validation of Various Phenotyping Algorithms for Diabetes Mellitus Using Data from Electronic Health Records

Santiago Esteban, Manuel Rodríguez Tablado, Francisco Peper, Yamila S. Mahumud, Ricardo I. Ricci, Karin Kopitowski, Sergio Terrasa

Family and Community Medicine Division, Hospital Italiano, Buenos Aires, Argentina

Abstract

Precision medicine requires extremely large samples. Electronic health records (EHR) are thought to be a cost-effective source of data for that purpose. Phenotyping algorithms help reduce classification errors, making EHR a more reliable source of information for research. Four algorithm development strategies for classifying patients according to their diabetes status (diabetic; non-diabetic; inconclusive) were tested (one codes-only algorithm; one boolean algorithm, four statistical learning algorithms and six stacked generalization meta-learners). The best performing algorithms within each strategy were tested on the validation set. The stacked generalization algorithm yielded the highest Kappa coefficient value in the validation set (0.95 95% CI 0.91, 0.98). The implementation of these algorithms allows for the exploitation of data from thousands of patients accurately, greatly reducing the costs of constructing retrospective cohorts for research.

Keywords:

Diabetes Mellitus, Algorithms, Precision Medicine

Introduction

Recently, the progression towards precision medicine [1] has sought the development of large databases, allowing assessment of the impact of risk factors or treatments in specific subpopulations. This is usually a problem for classical cohorts, given the difficulty of enrollment and follow-up of a large enough number of patients [5]. Even more difficult is the situation for developing countries, given the usual lack of funds for local research [3].

Electronic health records (EHR) have been proposed as a solution to these two costs problems [10].

Phenotyping algorithms allow, through the combination of different variables extracted from the EHR, classifying patients according to their particular phenotype [8; 11]. Ideally, these algorithms must be validated and a metric should be estimated (accuracy, sensitivity and specificity, coefficient Kappa, F1 score, positive and negative predictive values) of the ability to classify patients compared to a gold standard. This facilitates the classification of large numbers of patients without the intervention of a human.

Boolean or rule-based algorithms are a common strategy for developing these algorithms. A different approach is the development of learners based on statistical learning, such as logistic regression or more recent methods such as decision trees, neural networks or support vector machines. The

different families of algorithms explore the multidimensional space of data in different ways so it can be beneficial to combine them. One way to do this is through stacked generalization. This methodology, described by Wolpert [4] for classification problems and by Breiman [7] for regression problems, seeks to improve the predictive power of the individual algorithms by developing a meta-learner incorporating the predictions of all algorithms as input, combining them, and then issuing a final prediction.

Our objective is to compare the performance of different classification strategies (only using standardized problems, rules-based algorithms, statistical learning algorithms and stacked generalization), for the categorization of patients according to their diabetic status (diabetic, not diabetic and inconclusive; diabetes of any type) using information extracted from EHR.

Methods

Study population

Patient information was extracted from the EHR of the Hospital Italiano in Buenos Aires, Argentina.

In order to have a training and a validation dataset, two samples of patients from different years (2005-2015; total n = 2463) were extracted. The only inclusion criterion was age (≥ 40 <80 years old by 1/1/2005 and by 1/1/2015 for each sample). The sampling was carried out using simple randomization. The training set (2005) featured 1663 patients. The validation set (2015) represented roughly 33% of the total sample (n = 800).

Feature extraction

Six variables were extracted: No. of standardized problems related to Diabetes Mellitus (DM) (inpatient, outpatient and emergency department codes); No. of filled oral hypoglycaemic or insulin prescriptions; No. of outpatient fasting glucose (FG) measurements ≥ 126 mg/dl; No. of outpatient FG measurements < 126 mg/dl; No. of HbA1c measurements $\geq 6.5\%$; No. of HbA1c measurements $< 6.5\%$. These variables were also used in previous research [6; 12]. Oral glucose tolerance measurements were left out in order to avoid making a diagnosis of gestational diabetes. Random ≥ 200 mg/dl blood glucose was not considered since it was difficult to establish if it coincided with diabetic symptoms, as indicated by the criteria of the American Diabetes Association (ADA).

Manual chart review

Four researchers manually reviewed all records and classified patients, analyzing all available information in the EHR. Patients were classified as:

- **Diabetics:** The ADA criteria [2] to classify patients as diabetics were used. Also, patients whose records stated that they were diabetics were classified as such.
- **Not diabetic:** To be classified as a non-diabetic, patients must have at least one FG measurement below 126 mg/dl, and must not have any references in their records regarding being diabetic or fulfill any of the ADA's criteria for DM.
- **Inconclusive:** Patients without a reference in their EHR regarding their diabetic status, nor a normal FG measurement, were classified as inconclusive. Those who had a single FG value above 126 mg/dl without a subsequent confirmatory measurement were characterized in the same way.

The level of agreement among researchers was assessed using the Kappa coefficient with a value of 0.92 (95% CI: 0.84, 0.99).

Algorithm development and validation process

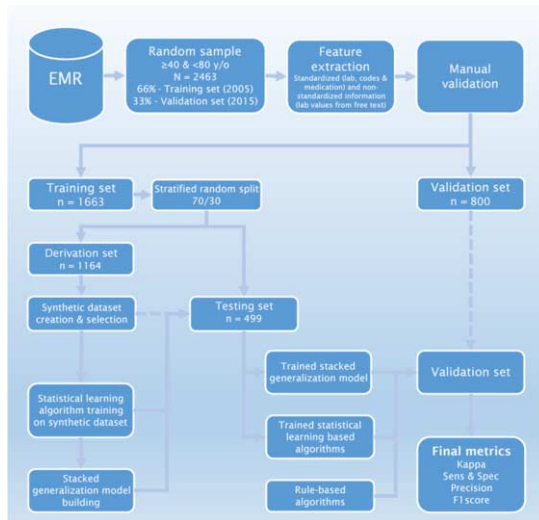


Figure 1 - General process of developing and validating phenotyping algorithms. EMR: Electronic Medical Records

Rules-based algorithms

In our study, we included two algorithms of this type:

- Classification of patients according to standardized codes: patients were classified as diabetics if they had at least one DM code in their EHR.
- Boolean logic algorithm (Adapted from Kho, Wilke, and Nichols): we used a combination of these three algorithms.

Algorithms based on statistical learning

Four of these learners were included individually: multinomial logistic regression, random forests, neural networks and support vector machines with radial kernel. According to their performance on the test set, the best one was evaluated in the validation set.

The problem of imbalanced datasets

We opted to use sampling techniques to adjust the imbalance between classes of the dependent variable. To select the best synthetic sampling algorithm, we divided the derivation dataset into a training and a test set. We then used the approach developed by Lopez et al. [9]: from the training set, we generated 19 sets of data by applying a combination of over and under sampling algorithms and analyzed them by means of four algorithms that use different approaches (neural networks, Elastic Net, Gradient Boosting Machine and C5.0). We then applied the trained learners on the test set (which remained unbalanced) and then ranked the datasets according to their performance. The best ranked dataset was used for the training of the statistical learning algorithms.

Development of the meta-learner

For the final prediction we selected the Elastic-Net algorithm. As a first step, we discarded those learners with significantly lower performance in the different subsets of the repeated cross-validation (RCV) (set 1). Then, four selection strategies were used: 1. We kept those algorithms whose performance in the different subsets of the RCV were less correlated (Pearson coefficient <0.75; set 2); 2. Using hierarchical clustering (Euclidean distance, complete method), learners were clustered according to their patterns of performance in the subsets of the RCV; the best within each cluster at different height levels were selected (sets 3, 4, 5); 3. We selected those with better performance within each family of algorithms (support vector machines, neural networks, decision trees, instance-based algorithms, algorithms, Bayesian, discriminant analysis, and linear models (set 6). Each of these versions of the meta-learner was evaluated on the test set and the most parsimonious and best performing learner was selected as the final model.

Validation

For the validation process, the different algorithms were evaluated in the validation set. The Kappa coefficient was used as the performance metric.

All analyses were performed using R (R Foundation for Statistical Computing, Vienna, Austria.) URL: <https://www.R-project.org>.

Results

Table 1 shows the characteristics of the patients included in both datasets. We can observe that patients from the sample of 2015 (validation dataset) generally have a greater number of measurements.

Table 1 - Characteristics of patients included in the training and validation datasets. DM: Diabetes Mellitus; FG: Fasting glucose.

	Training Set (2005)			Validation Set (2015)		
	Non - DM	DM	Undefined	Non - DM	DM	Undefined
n	1249	121	293	698	58	44
Number of DM-related codes (median [IQR])	0.00 [0.00, 0.00]	1.00 [1.00, 1.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	1.00 [1.00, 2.00]	0.00 [0.00, 0.00]
Number of DM-related prescriptions filled	0.00 [0.00, 0.00]	3.00 [0.00, 7.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	13.50 [1.00, 34.25]	0.00 [0.00, 0.00]
Number of FG ≥ 126 mg/dl	0.00 [0.00, 0.00]	2.00 [1.00, 4.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	4.50 [2.00, 9.00]	0.00 [0.00, 0.00]
Number of FG <126 mg/dl	2.00	2.00	0.00	7.00	7.50	0.00

	[1.00, 3.00]	[0.00, 4.00]	[0.00, 0.00]	[3.00, 11.00]	[4.00, 15.75]	[0.00, 0.00]
Number of HbA1c $\geq 6.5\%$	0.00	1.00	0.00	0.00	2.00	0.00
	[0.00, 0.00]	[0.00, 2.00]	[0.00, 0.00]	[0.00, 0.00]	[0.00, 7.00]	[0.00, 0.00]
Number of HbA1c $< 6.5\%$	0.00	1.00	0.00	0.00	3.00	0.00
	[0.00, 0.00]	[0.00, 2.00]	[0.00, 0.00]	[0.00, 0.00]	[1.00, 6.00]	[0.00, 0.00]
Age (mean (sd))	60.52 (11.48)	66.22 (9.09)	53.94 (10.64)	60.12 (11.15)	66.47 (9.89)	55.00 (10.07)

The results of all proposed algorithms are presented in figure 2 and table 2.

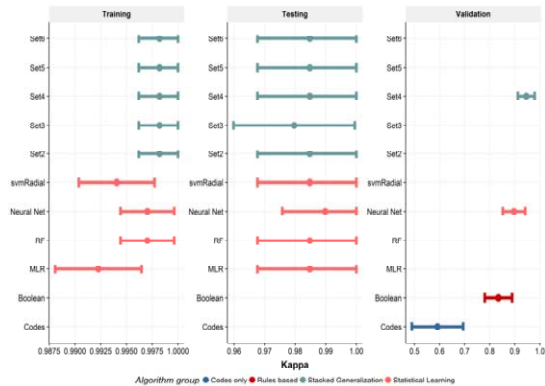


Figure 2 - Algorithm performance across training, test and validation sets.

Development and selection of the synthetic dataset

The dataset with best performance was a combination of Synthetic Minority Over-Sampling Technique (SMOTE) and Edited Nearest Neighbors (ENN) and was selected for the training of models. See figure 3.

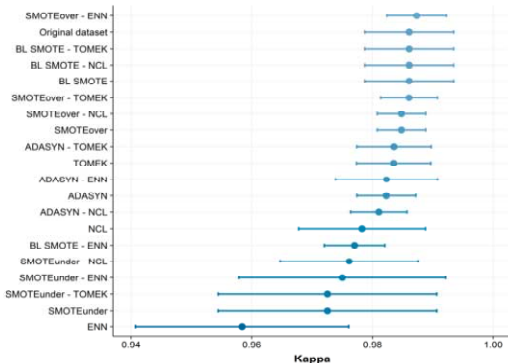


Figure 3 - Averaged performance of four algorithms trained on different datasets, in the testing set. SMOTE: Synthetic minority over-sampling technique; BL: Borderline; ADASYN: Adaptive synthetic sampling; over: over-sampling; under: undersampling; ENN: Edited nearest neighbors; Tomek: Tomek links; NCL: Neighborhood cleaning rule.

Algorithms based on statistical learning

The network neural presented the best performance in the test set and therefore was evaluated in the validations set (figure 2 and table 2).

Development of the meta-learner

Selection of models with the best performance

Figure 4 shows the performance of the 16 algorithms under consideration for building the meta-learner. Five algorithms showed Kappa coefficient values clearly below the rest and were excluded. The remaining eleven constitute set 1.

Selection of models based on low correlation

We assessed the level of correlation of performance of the algorithms in the different subsets of RCV (set 1). We detected those correlations greater than 0.75; then the average correlation of both algorithms were compared and the one with the highest mean correlations with all other models was removed. RRF, GBM, EGB - Linear and Elastic Net were removed. The remaining six formed set 2. The results are presented in figure and table 2.

Table 2 - Performance metrics for four algorithm strategies in the validation set. NDM: Non-Diabetes Mellitus; DM: Diabetes Mellitus; INC: Inconclusive

Perf. Measure	Confusion Matrix			Kappa	
	Pred. x Reference				
Codes		NDM	DM	INC	0.59 (0.49, 0.69)
	NDM	697	12	44	
	DM	1	46	0	
	INC	0	0	0	
Boolean		NDM	DM	INC	0.83 (0.78, 0.89)
	NDM	668	11	19	
	DM	0	58	0	
	INC	4	0	40	
Feedforward Neural Net		NDM	DM	INC	0.90 (0.85, 0.94)
	NDM	682	0	4	
	DM	16	58	0	
	INC	0	0	40	
Stacked Generalization (set 4)		NDM	DM	INC	0.95 (0.91, 0.98)
	NDM	693	1	4	
	DM	5	58	0	
	INC	0	0	40	

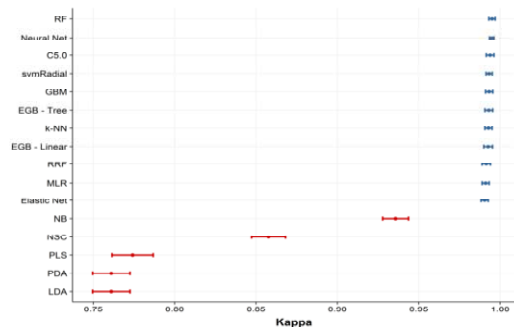


Figure 4 - Algorithm performance averaged over 50 cross-validation training subsets. RF: Random forests; svmRadial: Support vector machines with radial kernel; GBM: Gradient boosting machine; EGB: Extreme gradient boosting; k-NN: k nearest neighbors; RRF: Regularized random forests; MLR: Multinomial logistic regression; NB: Naïve bayes; NSC: Nearest shrunken centroids; PLS: Partial least squares; PDA: Partial discriminant analysis; LDA: Linear discriminant analysis.

Selection based on hierarchical clustering

Figure 5 shows the dendrogram generated from the hierarchical clustering of the performance of the different algorithms in the subsets of RCV. Three cutoff points were chosen. For cutoff level, the algorithm with best performance by cluster was selected (Set 3: RF, svmRadial, k-NN, GBM; Set 4: 3 Set + MLR; 5 set: Set 4 + Neural Net, EGB - Linear). Each set was then used as input for a different version of the Elastic Net-based meta-learner. Each version was tested in the test set. Set 4 presented the best combination of performance and parsimony and was selected to be applied to the validation set.

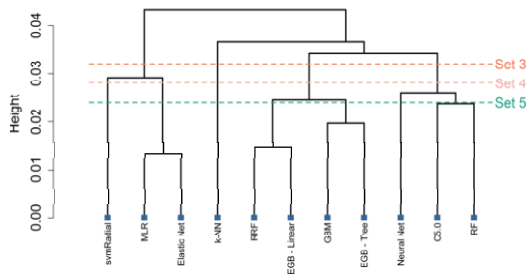


Figure 5 - Hierarchical clustering of classification algorithms based on their performance in the CV datasets.

Selection based on algorithm family

We selected those algorithms with the best performance within each family, to form set 6 (SVM radial Kernel, neural network with a single layer, EGB - Linear, Random Forests and k-Nearest Neighbors).

Selection of the best meta-learner

Finally we compared the performance of different versions of the meta-learner in the test set. The version that used set 4 as input presented the best combination of performance and parsimony. Its ability to classify patients was then evaluated in the validation set (figure 2 and table 2).

Discussion

Three of the four algorithms evaluated on the validation set showed excellent performance in terms of the Kappa coefficient. Our decision to use this metric above others more commonly used, such as accuracy, was related to the need for high levels of classification within each of the categories given the imbalance of classes in our dataset. This can be affected when the considered metric is accuracy, since it does not consider the agreement for each class but only the level of total error in the confusion matrix.

Each strategy presents pros and cons that are important to consider, since performance is not the only variable to take into account when selecting an algorithm to apply. Algorithms based on rules have the advantage of being simple and easily scalable with minimum processing time. However, we found that their performance is clearly lower than those based on statistical learning and stacked generalization. The neural network showed a high level of optimism (the difference in performance between the training and the validation set). This was less significant for the stacked generalization, which would go in hand with the main objective of this strategy — to reduce overfitting to the training set. The version of the meta-learner that used set 4 as input showed the best metrics of classification

in the validation set. Its implementation for research would be helpful, but probably not so for real-time applications given the higher processing time compared to simpler approaches.

Conclusion

We evaluated the performance of different strategies for the development of diabetes phenotyping algorithms using data extracted from an EHR from Argentina. The stacked generalization strategy showed the best metrics of classification in the validation set. The implementation of these algorithms enables the exploitation of the data of thousands of patients accurately, and a reduction of costs compared to traditional ways of collecting data for research. Thus, millions of patients from developing countries could benefit from local and specific data that could lead to treatments that take into account all their characteristics (genetic, environmental, habits, etc.) as it is the objective of precision medicine.

References

- [1] The Precision Medicine Initiative, in.
- [2] A.D. Association, Standards of Medical Care in Diabetes--2014, *Diabetes Care* **37** (2014), S14-S80.
- [3] A. Barceló and S. Rajpathak, Incidence and prevalence of diabetes mellitus in the Americas, *Revista Panamericana de Salud Publica* **10** (2001), 300-308.
- [4] W. DH, Stacked Generalization, *Neural Networks* **5** (1992), 18.
- [5] J.E. Fradkin, M.C. Hanlon, and G.P. Rodgers, NIH Precision Medicine Initiative: Implications for Diabetes Research, *Diabetes Care* **39** (2016), 1080-1084.
- [6] A.N. Kho, M.G. Hayes, L. Rasmussen-Torvik, J.A. Pacheco, W.K. Thompson, L.L. Armstrong, J.C. Denny, P.L. Peissig, A.W. Miller, W.-Q. Wei, S.J. Bielinski, C.G. Chute, C.L. Leiserson, G.P. Jarvik, D.R. Crosslin, C.S. Carlson, K.M. Newton, W.A. Wolf, R.L. Chisholm, and W.L. Lowe, Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study, *Journal of the American Medical Association: JAMA* **19** (2012), 212-218.
- [7] B. L., Stacked Regressions, *Machine Learning* **24** (1996), 15.
- [8] K.P. Liao, A.N. Ananthakrishnan, V. Kumar, Z. Xia, A. Cagan, V.S. Gainer, S. Goryachev, P. Chen, G.K. Savova, D. Agniel, S. Churchill, J. Lee, S.N. Murphy, R.M. Plenge, P. Szolovits, I. Kohane, S.Y. Shaw, E.W. Karlson, and T. Cai, Methods to Develop an Electronic Medical Record Phenotype Algorithm to Compare the Risk of Coronary Artery Disease across 3 Chronic Disease Cohorts, *PLoS One* **10** (2015), e0136651.
- [9] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics, *Information Sciences* **250** (2013), 113-141.
- [10] S.N. Murphy, G. Weber, M. Mendis, V. Gainer, H.C. Chueh, S. Churchill, and I. Kohane, Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2), *J Am Med Inform Assoc* **17** (2010), 124-130.
- [11] L.V. Rasmussen, W.K. Thompson, J.A. Pacheco, A.N. Kho, D.S. Carrell, J. Pathak, P.L. Peissig, G. Tromp, J.C. Denny, and J.B. Starren, Design patterns for the development of electronic health record-driven phenotype extraction algorithms, *J Biomed Inform* **51** (2014), 280-286.
- [12] R.A. Wilke, R.L. Berg, P. Peissig, T. Kitchner, B. Sijercic, C.A. McCarty, and D.J. McCarty, Use of an Electronic Medical Record for the Identification of Research Subjects with Diabetes Mellitus, *Clinical Medicine and Research* **5** (2007), 1-7.

Address for correspondence

Corresponding author: Dr. Santiago Esteban

E-mail : santiago.esteban@hospitalitaliano.org.ar;
santiagoesteban@gmail.com

Detecting Signals of Interactions Between Warfarin and Dietary Supplements in Electronic Health Records

Yadan Fan^a, Terrence J. Adam^{a,b}, Reed McEwan^c,
Serguei V. Pakhomov^{a,b}, Genevieve B. Melton^{a,d}, Rui Zhang^{a,d}

^aInstitute for Health Informatics, ^bCollege of Pharmacy, ^cAcademic Health Center-Information Systems,

^dDepartment of Surgery, University of Minnesota, Minneapolis, MN, USA

Abstract

Drug and supplement interactions (DSIs) have drawn widespread attention due to their potential to affect therapeutic response and adverse event risk. Electronic health records provide a valuable source where the signals of DSIs can be identified and characterized. We detected signals of interactions between warfarin and seven dietary supplements, viz., alfalfa, garlic, ginger, ginkgo, ginseng, St. John's Wort, and Vitamin E by analyzing structured clinical data and unstructured clinical notes from the University of Minnesota Clinical Data Repository. A machine learning-based natural language processing module was further developed to classify supplement use status and applied to filter out irrelevant clinical notes. Cox proportional hazards models were fitted, controlling for a set of confounding factors: age, gender, and Charlson Index of Comorbidity. There was a statistically significant association of warfarin concurrently used with supplements which can potentially increase the risk of adverse events, such as gastrointestinal bleeding.

Keywords:

Electronic Health Records; Natural Language Processing; Warfarin

Introduction

Drug and supplement interactions (DSIs) have drawn widespread attention in recent years due to the increased prevalence of dietary supplements worldwide. Patients often take prescribed medications along with dietary supplements to boost the immune system or to mitigate the side effects of a particular treatment. A major safety concern is the potential for dangerous adverse events caused by DSIs, particularly for drugs with narrow therapeutic indexes, such as warfarin. Increasing our knowledge base about DSIs will assist pharmacists and healthcare providers to provide guidance to patients on the safety and efficacy of the concomitant use of prescribed medications and dietary supplements, especially for the elderly, who have increased vulnerability to DSIs. Given the difficulty of testing DSIs in human populations, information on DSIs mostly comes from *in vitro*, animal research, or case reports [1]. Unfortunately, this information is under-reported and can be inconsistent. Also, clinical trials for drug approval may not reveal DSIs since supplements and drug-drug interactions often require large patient populations for adequate study power, especially with rare events. Our prior study identified several known and potential DSIs by mining 23 million biomedical literature abstracts (MEDLINE) [2]. Although the biomedical literature may help us to infer DSI knowledge and potential hypotheses for novel DSIs, we may also leverage

electronic health record (EHR) systems to complement DSI understanding and validate DSI hypotheses. EHRs offer a rich source of patient information since they serve as the primary patient care documentation platform for clinical care delivery. Some EHR data of interest to study DSIs include medication information, problem lists, laboratory data, and clinical notes.

Warfarin, as one of the most commonly prescribed anticoagulants, is widely used to treat and prevent thromboembolic events associated with atrial fibrillation, heart valve replacement, myocardial infarction and existing thromboembolic disease. However, warfarin is often involved in interactions with supplements because its metabolism involves multiple active metabolic pathways [3]. Natural products such as garlic, ginger and ginkgo are among the most common supplements implicated in DSIs with warfarin. Garlic has the side effect of platelet inhibition, which can increase the risk of bleeding when used with anticoagulant drugs [4]. Ginger can inhibit thromboxane synthetase and therefore lead to prolonged bleeding times [4]. Ginkgo will increase the International Normalized Ratio (INR) with warfarin, and ginseng might reduce the effect of warfarin [4]. Vitamin E can interact with warfarin due to blood thinning effects, especially in Vitamin K deficient individuals [5]. According to the Natural Medicines Comprehensive Database (NMCD) [6], warfarin also has significant interactions with alfalfa, grapefruit, and St. John's Wort.

In our previous study, we found that clinical notes contain some supplement mentions that do not exist in the medication list [7]. Much information about supplement use is embedded in clinical notes, and thus in this study we demonstrate that informatics techniques, especially natural language processing (NLP) methods, are effective in extracting supplement use status information from clinical notes. Specifically, we conducted survival analysis to test the significance of the concomitant use of warfarin and supplements associated with the appearance of adverse events based on the EHR data from University of Minnesota Clinical Data Repository (UMN-CDR). We focused our assessment on the adverse interactions of warfarin with seven dietary supplements: alfalfa, garlic, ginger, ginkgo, ginseng, Vitamin E, and St. John's Wort, with potential interactions indicated in the NMCD knowledge base. Adverse events were limited to embolic stroke and thromboembolism, which are defined as warfarin treatment failure events. In addition, bleeding (including GI bleeding) and the subset of patients with GI bleeding were evaluated and defined as side effects of warfarin in the context of this study.

Background

Warfarin potential interactions

Warfarin is a medication with a long history of clinical use due to its effect on the human coagulation system, but also as a poisoning agent/rodenticide due to these same pharmacologic characteristics. This dichotomy is carefully balanced clinically by having regular testing of the therapeutic response with the use of warfarin. Patient monitoring is managed by assessing blood coagulation with prothrombin times and INR testing which provide standard clotting measures. The testing is typically done on a monthly basis or even more frequently with substantial dosage changes or use of medications associated with drug-drug interactions. This intensive follow-up therapy helps to reduce the clinical risk of excess therapeutic effect (anticoagulation) which can result in bleeding. In addition, the monitoring allows dosages to be constantly adjusted to maintain adequate levels of anticoagulation to prevent thromboembolic events. Given this frequent follow-up by patients, the shifts in levels of anticoagulation are typically noticed before clinically significant events occur. However, if patients initiate new medications or supplements shortly after regular testing, they may be at risk for several weeks before routine testing can detect drug responses outside of the usual therapeutic range.

BioMedICUS and NLP-PIER

Both BioMedICUS and NLP-PIER are tools developed by the NLP/Information Extraction group at the UMN. BioMedICUS (BioMedical Information Collection and Understanding System) [8] is an open-source NLP system based on the Unstructured Information Management Architecture – Asynchronous Scaleout (UIMA-AS) architecture [9] specializing in NLP-related information extraction and understanding of clinical notes. NLP-PIER (Patient Information Extraction for Researchers) is a web-based search engine for clinical notes from the EHR [10]. Clinical notes in the CDR are run through a BioMedICUS NLP pipeline and indexed for use in NLP-PIER. BioMedICUS identifies UMLS Metathesaurus concepts (concept unique identifiers, or CUIs) from lexical variants expressed in the notes, and whether the identified concepts were used in a negated context. These negation-qualified CUIs are added to a set of 15 patient-related and encounter-related note attributes from the CDR, including five attributes from the HL7-LOINC document ontology [11]. Attributes and CUIs are stored in an Elasticsearch cluster along with the clinical note itself, which is run through an Elasticsearch snowball analyzer when it is indexed. This setup enables full text searches to be run on research-related note sets within NLP-PIER. Search terms can be expanded by specifying UMLS CUIs as part of the search query and results can be filtered using the attributes.

Methods

The method of this study consists of five steps: 1) data collection: search for patients taking warfarin and collect information about patients' demographics, warfarin usage, diagnosis and clinical notes; 2) NLP for supplement information extraction: apply NLP module to extract information about supplement use in clinical notes; 3) structured data query: query medication table for supplement use and diagnosis table for adverse events; 4) data combination: combine information from structured and unstructured data to generate a comprehensive data set for each patient; 5) statistical analysis: conduct survival analysis

to detect the significance of adverse events caused by concurrent use of warfarin and supplements.

Data collection

Patient cohort data in the Epic EHR were extracted from the UMN-CDR hosted by the Academic Health Center-Information Services (AHC-IS) exchange platform and supported through the Clinical Translational Science Institute (CTSI) at the UMN. The data in the CDR comes from the EHR of more than 2 million patients who sought health services at eight hospitals and over 40 clinics. Data are available for hospital visits starting from 2011. IRB approval was obtained for accessing the clinical notes.

Patients who have warfarin prescriptions from 2011 to September 2015 were included by using both generic name and brand names (*i.e.*, Coumadin, Panwarfin, Sofarin) of warfarin. Patients with medication records showing at least one warfarin prescription and complete information about the warfarin start date and end date were included in this study. The data from a total of 48,426 patients were stored in AHC-IS data shelter, which included patients' demographic information, diagnosis, and medications. Clinical comorbidities were calculated using the Charlson Index of Comorbidity. Their corresponding clinical notes were processed by BioMedICUS and indexed by NLP-PIER for further information extraction.

NLP for supplement information extraction

Since much of the information about supplement use was embedded in clinical notes, we retrieved the related clinical notes using PIER for further information extraction. Selected supplements and all their lexical variants were used for retrieving clinical notes. For instance, "ginkgo", "Vitamin E" and "St. John's Wort" have their lexical variants including "gingko", "ginko" and "ginkoba", "Vit E", "St. Johns Wort", "St. John Wort", "St John's wort", "St Johns wort", "St John Wort", respectively. However, we found instances of negative mentions (such as discontinuation of supplements) of supplements in the notes, such as "she may try ginkgo biloba for her memory issues" or "Denies using St John's Wort". Therefore, we applied a NLP module to classify the use status of the supplements, especially the active ones, such as "started" and "continuing", and also filtered the irrelevant clinical notes, such as "discontinuing" and negative mentions.

In our prior study [12], we used machine learning-based methods to automatically classify the use status of the supplements into four categories (Continuing (C), Discontinued (D), Started (S), Unclassified (U)). A total of 1,300 sentences on 25 most commonly consumed supplements were randomly selected and annotated. The training set consisting of 1000 sentences of 10 supplements was used to select the optimal algorithm with the identified feature sets. The test set included 300 sentences on the remaining 15 supplements. We trained four algorithms with seven different feature sets in the study. The best model (*i.e.*, Support Vector Machine (SVM) with the feature set of unigram, bigram and indicator words within window size of four tokens on both sides of supplement mention) achieved F-measure of 0.906, 0.913, 0.914, 0.715 for status C, D, S, U on the test set, respectively. We further applied the trained SVM model on the notes retrieved in this study. We only consider the "Continuing" and "Started" categories since they are the active status for supplement use. "Discontinued" category may hold important information about the past use of supplements, however, the start date of the supplements remains unclear, therefore, "Discontinued" was considered negative case in this study.

Structured data query

Supplements and warfarin: Both structured and unstructured data were used in the search for supplement use. Since some supplement products were also included in the medication tables, we queried them using the common names and the lexical variants of supplements as above. Warfarin information was also retrieved from the medication table.

Adverse events: A list of ICD9-CM codes (International Classification of Disease, Ninth Revision, Clinical Modification) was used to identify patients having adverse events, which included gastrointestinal (GI) bleeding, general bleeding (GI bleeding included), embolic stroke and thromboembolism (see Table 1).

Table 1—ICD-9 Codes for Adverse Events

Diagnosis	ICD-9 Codes
GI bleeding	530.7, 530.82, 531.2, 531.4, 531.6, 532.2, 532.4, 532.6, 533.2, 533.4, 533.6, 534.2, 534.4, 534.6, 535.x1, 537.83, 562.02, 562.03, 562.12, 562.13, 569.3, 578.x
general bleeding	964.2, 964.5, E934.2, E934.5, 459.0, 285.1, 286.59, 362.81, 596.7, 599.70, 599.71, 719.1x, 782.7, 784.7, 784.8, 786.30, 786.39, 423.0, 423.1, 423.9, 568.81, 530.7, 530.82, 531.2, 531.4, 531.6, 532.2, 532.4, 532.6, 533.2, 533.4, 533.6, 534.2, 534.4, 534.6, 535.x1, 537.83, 562.02, 562.03, 562.12, 562.13, 569.3,e 578.x, 430, 431
embolic stroke	346.6x, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434, 434.01, 434.1, 434.10, 434.11, 434.9, 434.90, 434.91, 436
thromboembolism	451.1x, 453.4x, 453.5x, 453.8, 453.9, 415.1x

Data combination

Structured and unstructured data were combined to generate a comprehensive data set for each patient. Patients were further divided into two groups based on their use of seven supplements. The supplements-reported group include patients who have at least one prescription record showing that they take warfarin and at least one of the seven supplements concurrently. The warfarin-only group include patients taking warfarin only and do not have exposure to any of the seven supplements during the time period six months before the warfarin initiation to the first occurrence of an adverse event of interest based on the EHR data. We used the note date as the supplement's start date if the start date was not specifically mentioned in the clinical notes and medication table. For example, "Pt has started taking ginseng" and "Patient has been taking garlic" indicate that the patient has already started taking the supplements before the visit; however, detailed information about the start date is unavailable in the notes. Additionally, it usually takes weeks for warfarin to demonstrate a stable therapeutic level. The initial titration phase may be a time of increased risk of adverse events until a stable warfarin dose is reached. To reduce the drug titration effect bias, the first 30 days of warfarin use were eliminated for both groups [13]. For the supplements-reported group, day

1 was the first day when any of the supplements were first noted in the EHR after eliminating the first 30 days of warfarin use. For the warfarin-only group, day 1 was actually the day 31 for the warfarin use.

Survival analysis

Cox proportional hazards models were fitted to compare the hazard of adverse events between two groups, controlling for a set of confounding factors including age, gender, and comorbidities. All the patients were followed for one year for the first occurrence of adverse events. Follow-up ended with the first adverse event, or the end of the warfarin therapy. Kaplan-Meier survival curves were also plotted.

Results

A total of 41,257 patients were included in the study, among which 2,640 subjects were in the supplements-reported group who took warfarin and at least one of the seven supplements concurrently. The control group included 38,617 subjects in the warfarin-only exposures.

The number of patients taking each of the seven supplements were counted based on the information from both structured and unstructured data. The results in Table 2 indicate that the identification of supplement use was much larger with the use of the combination of structured and unstructured data approach, especially for garlic and ginger, since much of the information about dietary supplements related to food such as "garlic bread" and "ginger tea" were detected by our NLP module.

The hazard ratio, 95% CI, and p-value for the four adverse events are listed in Table 3. The results show that the hazard ratio of the four adverse events in the supplements-reported group are statistically significant and higher in the supplement exposed patients when compared with the warfarin-only group. The results indicate taking warfarin concurrently with supplements is associated with side effects such as bleeding, or therapeutic failure events like embolic stroke.

The Kaplan-Meier survival curves for four adverse events were shown in Figure 1. The results of the log-rank test indicate the survival curves for the supplements-reported group and the warfarin-only group are significantly different ($P < 0.01$) in GI bleeding, general bleeding, and embolic stroke, however, for thromboembolism, there is no significant difference in the curves between the supplements-reported

Table 2—The Number of Patients with Mention of Supplements Use in Structured and Unstructured Data

Supplements	Structured	Structured and Unstructured
alfalfa	30	68
garlic	329	925
ginger	100	1296
ginkgo	141	276
ginseng	42	109
Vitamin E	2273	4145
St. John's Wort	22	44

Table 3– Multivariable Cox Proportional Hazards Regression for Adverse Events (supplements-reported group VS. warfarin-only group)

Adverse events	HR (95% CI)	p-value
GI Bleeding	1.30 (1.08, 1.57)	0.005
General Bleeding	1.20 (1.07, 1.34)	0.002
Embolic Stroke	1.27 (1.06, 1.51)	0.008
Thromboembolism	1.13 (1.02, 1.25)	0.021

group and warfarin-only group.

Discussion

The literature has shown there may be potential adverse interactions between warfarin and supplements, however, in many cases, the limited available data impedes the assessment of the potential risk associated with concurrent use. Additionally, the data is hard to detect in clinical trials due to limited sample sizes and high costs for the evaluations.

Due to its blood thinning effect, patients on warfarin are warned to be careful taking other supplements, such as ginkgo, ginger, Vitamin E, which can potentially increase the risk of bleeding events. Our confirmation of these potential adverse interactions provides evidence to support the current clinical guidance and provides data to assess drug safety with DSIs. For example, alfalfa contains a large amount of Vitamin K, which can reduce the anticoagulant activity of warfarin [14]. Alfalfa was part of the original research on Vitamin K metabolism and was one of the first substances on which Vitamin K was synthesized. This finding is consistent with the expected response with Vitamin K directly reversing the effects of warfarin. Taking St. John's Wort induces cytochrome P450 2C19 which may clinically affect warfarin [15]. Ginkgo also affects the CYP3A4 path by inducing the enzyme which may affect the R-enantiomer of warfarin [16].

This study demonstrates the feasibility of using clinical data from EHR to detect the signals for adverse events associated with drug and supplement interactions. The results of the study as noted in the hazard ratio indicate a higher risk of adverse events and therapeutic failure well beyond typical screening triggers to assess the signal for potential adverse events. Additional assessment of the clinical cases will be needed to confirm the temporal and pharmacological patterns with the results to better assess the risk of exposure to supplements. However, from a medication safety perspective, the approach substantially reduces the assessment effort by patient safety officers or clinician managers to identify potential drug safety issues.

Though we used structured data for our outcomes assessment, our results provide support for the use of unstructured data to assess clinical exposures and outcomes. In addition, the combination of structured data (i.e., structured medication table) with unstructured data (i.e., clinical notes) in identifying supplements use has shown that clinical notes contain valuable information related to supplements which can complement structured data for DSI detection in the EHR. It is noted from our study that very little information about supplements is stored in the medication table since dietary supplements are regulated as food and can be obtained over the counter without

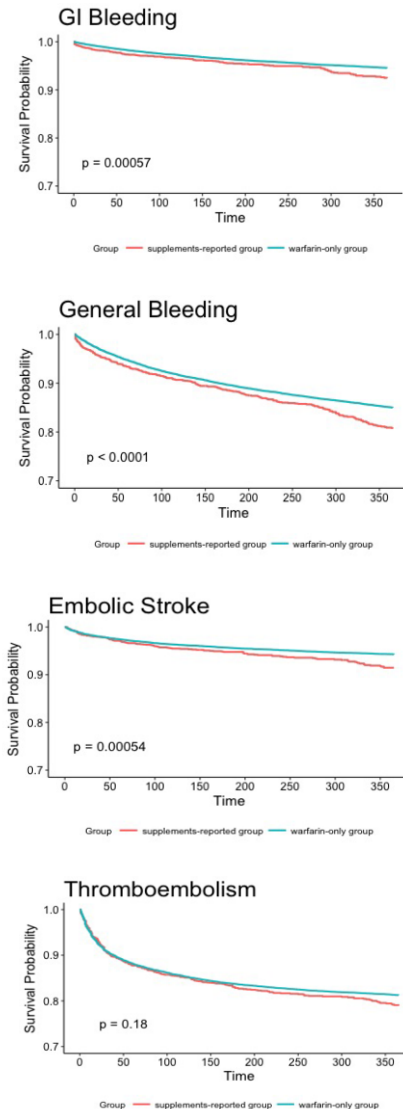


Figure 1– Kaplan-Meier Curves for Adverse Events

a prescription, consequently, much of the information about supplement consumption is documented in clinical notes during the medical encounter. Therefore, the combination of text information with a structured medication module is necessary for the supplements use identification, where NLP is essential for extracting supplements use related information from clinical notes.

One limitation of the study is that for some patients, the actual start date of supplements is before the date of the clinical note because we found some patterns like 'she has started ginger two months ago', which could lead to misclassification of the exposure in the statistical analysis. The data has limitations on both the medication and supplement orders, which limits the ability to directly assess the association between exposures and clinical outcomes. Correlation of these results with other peripheral data sources such as retail data, if available, could help better identify the acquisition of supplements for presumed use by patients. Patient diaries, medication adherence apps and other sources could also be considered as part of usual care processes to better identify supplement use.

The second limitation is that we only applied a limited set of common names of supplements in the search for notes related to supplement use which may have missed some supplement information. The recall of supplement information might be increased when using more complete supplement terms.

Another limitation of the study is that we did not take into consideration the end date of the supplements since this information was often missing. Such information may also be contained in the clinical notes but requires additional analysis. Future work including the development of the NLP system to accurately extract temporal supplement information from the clinical notes could better assess the relationships of supplement exposures to medication use and clinical outcomes.

Conclusion

This study indicates that it is possible to use existing EHR data to detect signals of DSIs. The current findings also demonstrate the feasibility of applying NLP methods to extract supplement usage information from clinical notes. Furthermore, these methods can likely be extended to detect other potential drug and supplement interactions providing an important approach for post-market surveillance for DSI as well as drug-drug interactions.

Acknowledgements

The project was partly supported by the National Center for Complementary & Integrative Health Award (R01AT009457) (Zhang), the Agency for Healthcare Research & Quality grant (#1R01HS022085-01) (Melton), the University of Minnesota Clinical and Translational Science Award (8UL1TR000114) (Blazer), and the UMN grant-in-aid award (Zhang). The clinical data was provided by the University of Minnesota's Clinical Translational Science Institute (CTSI) Informatics Consulting Service. The authors thank Fairview Health Services for their data support of this research.

References

- [1] I. Levy, S. Attias, E. Ben-Arye, L. Goldstein, and E. Schiff, Adverse events associated with interactions with dietary and herbal supplements among inpatients, *Br J Clin Pharmacol* (2016).
- [2] Zhang R, Simon G, Cairelli MJ, Rindfleisch T, Pakhomov S, Melton GB, Mining Biomedical Literature to Explore Interactions between Cancer Drugs and Dietary Supplements, *AMIA Jt Summits Transl Sci Proc* 2015 (2015), 69-73.
- [3] L.P.J. Vaes and P.A. Chyka, Interactions of warfarin with garlic, ginger, ginkgo, or ginseng: Nature of the evidence, *Annals of Pharmacotherapy* 34 (2000), 1478-1482.
- [4] C. Ulbricht, W. Chao, D. Costa, E. Rusie-Seamon, W. Weissner, and J. Woods, Clinical Evidence of Herb-Drug Interactions: A Systematic Review by the Natural Standard Research Collaboration, *Current Drug Metabolism* 9 (2008), 1063-1120.
- [5] M. Podszun and J. Frank, Vitamin E-drug interactions: molecular basis and clinical relevance, *Nutr Res Rev* 27 (2014), 215-231.
- [6] Natural Medicines Comprehensive Databases [Internet]. <http://naturaldatabase.therapeuticresearch.com>. Accessed: December 5, 2016.
- [7] Zhang R, Manohar N, Arsoniadis E, Wang Y, Adam TJ, Pakhomov S, Melton GB, Evaluating Term Coverage of Herbal and Dietary Supplements in Electronic Health Records, *AMIA Annu Symp Proc* 2015 (2015), 1361-1370.
- [8] BioMediCUS [Internet]. Available from: <https://github.com/nlpie/biomedicus>. Accessed September 20, 2015.
- [9] UIMA [Internet]. Available from: <https://uima.apache.org/>. Accessed September 20, 2015.
- [10] NLP-PIER [Internet]. <http://athena.ahc.umn.edu/nlppier/>. Accessed September 20, 2015.
- [11] McEwan R, Melton GB, Knoll BC, Wang Y, Hultman G, Dale JL, Meyer T, Pakhomov S. NLP-PIER: A Scalable Natural Language Processing, Indexing, and Searching Architecture for Clinical Notes. *AMIA Joint Summits Proceedings*. 2016. 150-9.
- [12] Fan YD, He L, Pakhomov S, Melton GB, Zhang R, Classifying Supplement Use Status in Clinical Notes, *AMIA Jt Summits Transl Sci Proc* 2017 (2017), 493-501.
- [13] Stoddard GJ, Shane-McWhorter L, et al., Ginkgo and Warfarin Interaction in a Large Veterans Administration Population, *AMIA Annu Symp Proc* 2015, 1174-1183.
- [14] Newall CA, Anderson LA, Philpson JD. *Herbal Medicine: A Guide for Healthcare Professionals*. London, UK: The Pharmaceutical Press, 1996.
- [15] Wang LS, Zhu B, Abd El-Aty A, et al. The influence of St. John's wort on CYP2C19 activity with respect to genotype. *J Clin Pharmacol* 2004;44:577-81.
- [16] Deng Y, Bi HC, Zhao LZ, Wang XD, Chen J, Ou ZM, Ding L, Xu LJ, Guan S, Chen X, Zhou SF, and Huang M Induction of Cytochrome P450 3A by the Ginkgo biloba Extract and Bilobalides in Human and Rat Primary Hepatocytes, *Drug Metab Lett*. 2, 6.

Address for correspondence

Rui Zhang, PhD, University of Minnesota, Minneapolis, MN, USA
Email: zhan1386

From Bench to Bedside: A View on Bioinformatics Pipelines

Blanca Flores^a, Dirk Hose^b, Anja Seckinger^b, Petra Knaup^a, Matthias Ganzinger^a

^a Institute of Medical Biometry and Informatics, Heidelberg University, Heidelberg, Germany,

^b Department of Internal Medicine, Heidelberg University Hospital, Heidelberg, Germany

Abstract

Although sequencing technology has become widely available in recent years, the steps in bioinformatics pipelines are time-consuming and barely standardized. New tools to improve individual steps in a pipeline are frequently published and configurations can be quickly adapted to use new versions. We performed case studies with a representative set of pipeline management tools using the GEP-R pipeline, and a qualitative study of different software packages covering relevant classes of software tools. We use a software toolset of R environment, Docker, KNIME, and BPEL to review our first aim of technical and organizational challenges. We propose snapshotting, documentation management, and a hybrid approach for our second aim of approaches to reproducibility. In order to have fully reproducible results derived from raw data, we think that it is necessary to archive biomedical analysis pipelines and their necessary software components.

Keywords:

Medical Informatics; Computational Biology; Software

Introduction

Sequencing and analysis of genomic data have become valuable technologies in both research and clinical settings. An effective and rapid analysis of data can support the identification of target genes and decision-making on diagnosis and treatment options [1]. Sequencing technology has become relatively cheap and widely available. In contrast, the steps of bioinformatics analysis following sequencing are still time-consuming and barely standardized. Typically, bioinformatics facilities develop and maintain their own chain of analytical steps, also known as an analytical pipeline. A variety of computer programs are available for each step. They have specific characteristics that make them more or less suitable for an analysis target. Each program used in each analytical step generates output data that can be used as input for the next program in the analytical pipeline. Since new tools for improving individual steps of a pipeline are frequently published, pipeline configurations can be quickly adapted to make use of the new versions.

Management of bioinformatics pipelines is facilitated by tools used for tasks such as the execution of serial and parallel steps and handling complex dependencies, varied software, data file types, and fixed and user-defined parameters [2]. Selection of tools in the pipeline depends on user requirement and implementation setting. Unfortunately, newer versions of packages can lead to unintended side effects in the analytical pipeline that may cause different results in calculations. Even small changes, like handling decimal places differently, can eventually lead to changes in classification of data or decision-making.

The problem of changing pipeline components is intensified by modular software packages such as the statistical software R [3]. Typically, a certain function is encapsulated into a package. However, packages often make use of functions provided by other packages, with each package having its own update cycle. As a consequence, complex packages have a complex dependency tree and are common in bioinformatics.

The differences in the pipeline implementation and frequent changes in tools and their versions affect the system's stability, reproducibility, and validation. For example, a change in risk-classification software as mentioned above might lead to attribution to a different prognostic group for a patient. In such a case, problems occur when results are compared to those of other patients classified on different versions or if procedures are repeated on different versions. Consequently, technical and organizational steps are necessary to manage documentation, ensure version control, and validate pipeline configurations in a clinical setting.

Standardization and validation of the software tools are strong requirements for pipelines used in clinical implementation [4]. Thus, we propose two approaches for achieving reproducible pipelines: Snapshotting complete configurations, and documenting pipelines in detail.

Snapshotting

Taking a snapshot of a pipeline can be compared to freezing the whole runtime environment including all software components as they are installed at the time of taking the snapshot. The snapshot can be conserved for later reactivation or be transferred to another environment to replicate the pipeline. The environment can be built in a very interactive way, which makes it easy to implement even complex dependencies with heterogeneous version requirements. For example, the latest version of a certain package might depend on an older version of another package and hence it may not be possible to simply use the latest versions of all packages.

Pipeline documentation

Another way to conserve the pipeline configuration is to document the precise pipeline configuration. Preferably, this is done in a machine-readable way, so the pipeline can be rebuilt automatically. This approach relies on the software management capabilities of the programs used in the pipeline: they should support unattended installation methods in order to install the correct programs reliably. For modular software like R, functions have to be provided to extract the list of installed packages with versions and automatically reinstall the software from such a list.

A bioinformatics pipeline has been developed for a project called "clinically applicable, omics-based assessment of survival, side effects, and targets in multiple myeloma (CLIOMMICS)" during the research stage. It has further ma-

tured for use in clinical routine care. Multiple myeloma is a malignant disease characterized by molecular heterogeneity that can be assessed by gene expression profiling (GEP). The GEP-R is a reporting tool developed using R [3] and Bioconductor [5] and generates a medical letter summarizing the scores derived from the GEP analysis. Thus, it allows the automated interpretation of Affymetrix gene expression microarray profiles by using a bioinformatics pipeline. It aims to provide quality controlled, validated, and clinically digestible information that includes molecular classification, risk stratification, and assessment of target gene expression [2]. The result of the GEP-R bioinformatics pipeline is a report in the form of a letter to the physician in charge of patient treatment.

The aim of this paper is to describe the technical and organizational challenges of handling changes for bioinformatics pipelines, using the GEP-R project as an example. In addition, we propose two approaches to improve stability and reproducibility of the pipeline, as well as validity of its tools.

Methods

To investigate the possibilities for increasing reproducibility of bioinformatics pipelines we performed case studies with a representative set of pipeline management tools. We selected the software tools with the intention to cover the reproducibility strategies of *pipeline documentation* and *pipeline snapshot*. We do not aim to provide a comprehensive overview of pipeline management software found in Leipzig, and Curcin et al. [6, 7]; instead, we performed a qualitative study of different software packages covering relevant classes of software tools. These classes are:

- Package-based analysis software
- Virtualization software
- Graphical pipeline management software
- Document-oriented pipeline management software

We used the GEP-R pipeline as a test case.

GEP-R pipeline

The biomedical objectives of this pipeline are described in the original publication by Meissner et al. [2]. GEP-R is a typical pipeline for the analysis of DNA microarray data from a technical perspective. Specifically, .CEL files derived from Affymetrix U133 Plus 2.0 DNA microarrays are analyzed to classify stage, prognosis, and treatment options for patients with multiple myeloma. The pipeline consists of a series of Bioconductor functions to preprocess the raw microarray data and subsequently calculate relevant scores. In addition, quality control indicators are calculated. All results are combined and visualized in a format that can easily be interpreted by physicians by using the Business Intelligence and Reporting Tools (BIRT) framework [8]. The pipeline uses standard packages from R and Bioconductor repositories, and in addition, a modified non-standard version of a specific package is required. A workflow representation of the pipeline components is shown in Figure 1.

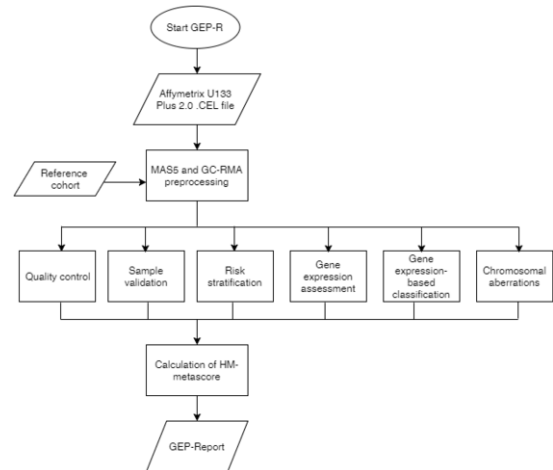


Figure 1- GEP-R pipeline

Software reviewed

Since R (often in combination with Bioconductor) is a very common pipeline component and at the same time a very complex system with complex internal dependencies, we consider it as a mandatory element of our test cases. Microsoft R repository is an approach for installing historic versions of R packages. It is possible to choose any date beginning 2014-09-17 within this repository and re-install a consistent R system based on package versions that were current on a specific date.

We chose KNIME Server version 4.3.2 [9] for a graphical pipeline management system for our test case. This tool has capabilities to interact with R and other external tools. The pipeline is manipulated as workflows via a graphical user interface. Workflows can be versioned and stored in a central repository.

We used Docker as a generic tool for preserving specific configurations of dependent software. With Docker, configurations can also be versioned and stored in repositories. In contrast to a pipeline management tool like KNIME, Docker cannot be used to provide a workflow by itself.

Finally, we investigated a document-oriented workflow management approach. BPEL can be used to design a workflow within the pipeline by describing its components with XML and Web Services. Changes in the pipeline can be tracked with version control and provide support for documentation management.

We performed our case studies on the basis of the bioinformatics pipeline established for the GEP-R report [2].

Results

The results are structured according to the research aims of our study. First, we describe the results for the evaluation of technical and organizational challenges. Second, we show our approaches for achieving reproducible pipelines by combining appropriate software components.

Aim 1: Technical and organizational challenges

R environment

R is a statistical software package that is very popular in biomedical research. However, its full scope cannot be leveraged by the monolithic core software, but relies on a huge amount of additional packages. R packages are typically developed by independent programmers and can be shared via the CRAN

network. As a result, release cycles of packages differ greatly and are typically not synchronized with the release cycle of core R. Since the behavior of packages can change during updates without notice, a newer version might break the pipeline.

The challenge of changing packages is attenuated to some extent by the R package checkpoint. This package was developed by Microsoft and allows rebuilding a consistent R environment using the Microsoft R Application Network (MRAN) repository. Unfortunately, it is not always the case that all packages used for a bioinformatics pipeline are at their current versions, since individual packages might have been added during the development process without updating the remaining packages.

In addition, bioinformatics applications are typically based on the Bioconductor tools. Bioconductor consists of more than 1200 R packages provided via its own repository system Bioc. Since Bioc is not included into the MRAN snapshots, it can be challenging if older versions of Bioconductor packages have to be installed manually.

Docker

Docker is an open source virtualization software that is installed on top of a variety of computer operating systems. It aims to virtualize individual applications in contrast to common virtualization approaches that work on complete computers. Docker provides mechanisms for easily moving the virtualized application (VA) between physical computers. The configuration of the VA can be versioned using the built-in snapshotting tool.

We successfully configured the complete R including Bioconductor and custom R packages inside a Docker container. This container can be copied to any computer that is intended for running our pipeline. Since the VA includes all required packages, we ensure consistent pipeline results across all instances. The integrity of the container is verified by comparing checkpoints of the VA's snapshot.

Graphical pipeline management system (KNIME)

KNIME is a general-purpose data analytics tool that provides a user interface for preparing a specific pipeline by graphically combining predefined components called nodes. For many analytical steps that typically occur in bioinformatics pipelines, nodes are shipped with KNIME or can be installed from an additional repository. However, not all specialized functions of Bioconductor are natively available in KNIME. Access to Bioconductor functions is available via KNIME's special R-nodes. These nodes are used for seamless integration of R-programs into KNIME workflows. Thus, data can be preprocessed in KNIME and R functions can be invoked as necessary within a workflow. The KNIME Server that we used for this test includes repository for storing and versioning pipeline configurations.

Document oriented pipeline management (BPEL)

The Business Process Execution Language for Web Services known as BPEL [10] is a process execution language standardized by OASIS. It allows the design of workflows using Web Services and represents data with XML specifications. Workflows can be linked and invoked by other Web Services, while inputs and outputs are assigned to variables to store data. Tools and technologies can be developed using Web Services to increase the level of automation in a process. Control structures used to manage tasks use either constructs that implement conditional branching and looping or activity containers to schedule sequential or parallel tasks. The capabilities of BPEL can be used for documentation management by

describing the workflow of a pipeline with XML and tracking changes with its support for version control.

The GEP-R pipeline components can be described by expressing the exchanged data between the analytical steps as XML forms and by defining the order in which the Web Services are invoked. The steps in a process are known as activities and different methods can be used to manipulate data in the pipeline, such as invoking services, initializing variables, assigning values, and performing calculations. These can be combined into algorithms to perform complex processes. Changes in the pipeline can evolve and different versions can be deployed depending on the specific needs by using version control. Support for flexibility and adaptability to changes mitigate technical and organizational challenges and also simplify the process of documentation management.

Aim 2: Approaches for reproducibility

Snapshotting

Snapshotting seems to be the approach that is easier to implement since it is usually not disruptive to established development and implementation procedures. The granularity of scope that is covered by a snapshot depends on the software tools that are used. One approach is to cover complex components like R with Bioconductor. We established a Docker container that contained all specific configurations for our GEP-R pipeline. Since we were able to use the common interactive mode of installation, setting up the image for this container was relatively easy. Since the installation of packages is not limited to repositories, it was possible to compile and install a specifically modified R package into the container. When problems occurred during the installation process, it was possible to fix them without reinstalling the container as a whole. Since the latest snapshot covers any changes that were made to the container, it is hard to forget to document modifications as long as the snapshot is generated. Docker supports the development process by providing a versioning system that allows for multiple versions and even forks of images. Containers can be moved or copied to other computer systems or even organizations in order to reproduce a specific pipeline.

While it is easy to manage pipelines as a whole using snapshots, it can be quite complex for users to understand what is happening inside such a snapshot. For understanding the process, it is disadvantageous that the generation of the snapshot was done in an interactive way without enforcing to log all steps and programs that were involved. Essentially, a snapshot represents a black box with a behavior that can be reproduced very well but that can be complex to comprehend.

Documentation management

Documenting all pipeline steps and components leads to a description of the pipeline that makes it easier to understand its functionality. Placing the individual steps of a pipeline into a process description—for example using BPEL—can be achieved with reasonable effort. The situation with complex components or subsystems like R is more problematic. While it is possible to extract a list of all packages installed including the respective version, this list cannot be easily used for installing a new instance of R that exactly duplicates the first installation. Typically, only the latest versions available in the corresponding repositories of R and Bioconductor are used. For the installation of custom versions of R packages, individual installation procedures are required that go beyond documenting only package name and version.

For a fully automatic replication of a pipeline and its runtime environment, it is further necessary to use a documentation scheme that can be interpreted by a computer. Extensive testing of the document is necessary in order to ensure a complete

and error-free system. For each test, the whole pipeline environment has to be rebuilt from scratch. The whole pipeline development process has to be adapted to the documentation approach as a result.

Hybrid approach

We also analyzed a hybrid approach combining snapshotting and pipeline documentation for our tests based on the GEP-R pipeline. We implemented a Docker container only for the specific R environment in this case. The pipeline itself was managed and documented using the KNIME Server. The functional steps of the pipeline can easily be followed in the graphical notation with this approach. The black-box-effect of the Docker container is less significant since it is clear which specific functional step is called in the R environment.

Discussion

We assessed software tools and two different strategies for the documentation and archival of bioinformatics pipelines in order to facilitate reproducibility in our study. There are several factors that may influence the quality of bioinformatics pipelines: availability of a variety of tools for different steps, ability to handle changes in user requirements or tool versions. These factors can cause side effects such as small differences in results, but may have bigger impacts in terms of classifications and decision-making.

We considered three characteristics that are important for bioinformatics pipelines: stability, reproducibility, and ability to validate tools and versions. Changes in requirements and configurations may occur due to different reasons, such as frequent updates of tools, availability of new tools, and changes in requirements. A proper documentation strategy and version tracking allow smoother transitions and facilitate changes without disrupting the analytical steps or results.

We described two approaches for achieving these characteristics, each having its advantages and disadvantages. Snapshotting the whole pipeline is relatively easy to implement using virtualization technologies like Docker. However, it is not easy to follow the pipeline's functionality. It is also not known for how long container runtime environment will be available.

Documentation-based approaches provide pipelines that can easily be understood when looking at the whole process. However, they are more complex to implement. On the component level, they also face the problem that specific software components might not be available in the future.

A compromise of these two approaches is to document the high-level pipeline using tools like KNIME or BPML. However, we suggest to use methods like Docker for individual components to conserve possible complex configurations for future use.

Research has been done to define formats for data expected to be used in future. More research is necessary to ensure the long-term availability of execution environments for the conserved pipelines.

Conclusion

Archiving raw data is considered good clinical practice as well as good research practice. However, in order for results derived from such data to be fully reproducible, it is necessary to archive biomedical analysis pipelines along with data as well as consider important requirements such as standardization and validation of all necessary software components.

Acknowledgements

CLIOMMICS is funded by the German Ministry of Education and Research within the e:Med initiative.
Grant id: 01ZX1609A.

References

- [1] A. Valencia, M. Hidalgo, Getting personalized cancer genome analysis into the clinic: the challenges in bioinformatics. *Genome Med* **4** (2012), 61
- [2] T. Meissner, A. Seckinger, T. Reme, T. Hielscher, T. Mohler, K. Neben et al., Gene expression profiling in multiple myeloma—reporting of entities, risk, and targets in clinical routine. *Clin Cancer Res* **17** (2011), 7240–7247
- [3] R Development Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2008
- [4] ISO/IEC, Health software –Part 1: General requirements for product safety. Geneva, IEC, 2016. (vol 35.240.80) 2016. Available from: URL:http://www.iso.org/iso/catalogue/catalogue_tc/catalogue_detail.htm?csnumber=59543 [cited 2016 Dec 16]
- [5] R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5** (2004), R80
- [6] J. Leipzig, A review of bioinformatic pipeline frameworks. *Brief Bioinform* (2016)
- [7] V. Curcin, M. Ghanem, Scientific workflow systems - can one size fit all? In: 2008 Cairo International Biomedical Engineering Conference (CIBEC), p. 1–9
- [8] D. Peh, N. Hague, J. Tatchell, *BIRT: A field guide. - Includes index*. 3rd ed., Addison-Wesley. Upper Saddle River, N.J., 2011. (Eclipse series)
- [9] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Köster, T. Meinl et al., KNIME - the Konstanz information miner. *SIGKDD Explor* **11** (2009), 26
- [10] Business Process Execution Language for Web Services., OASIS, 2007. Available from: URL:<http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.html>

Address for correspondence

Matthias Ganzinger
Institute of Medical Biometry and Informatics
Im Neuenheimer Feld 130.3
69120 Heidelberg
Germany
Email: Matthias.ganzinger@med.uni-heidelberg.de

A Personalized Adviser-Interpreter System for Monitoring Diabetic Patients' Weight Loss

Azadeh Kamel Ghalibaf^a, Seyed Mahmood Tara^b

^a Student Research Committee, Faculty of medicine,
Mashhad University of Medical Sciences, Mashhad, Iran

^b Department of Medical Informatics, Faculty of Medicine,
Mashhad University of Medical Sciences, Mashhad, Iran

Abstract

Diabetes and obesity are chronic diseases that need continuous follow-up. Analysis of the large volumes of patient data collected over time is time consuming and effortful for the physician. Automatic data abstraction and providing a summary of important events in the patient's follow-up history saves the physician's time and helps in making more informed decisions. The purpose of this study was to design an interpreter system that generates a tailored textual report about patients' adherence to a therapeutic regimen. The system has been designed in three phases: determining report structure, interpreting data, and generating textual-graphical reports for the patient and physician. The output report was evaluated by 12 diabetes nutrition specialists and eight diabetic patients. We used a questionnaire and a semi-structured interview approach to assess the quality of the report in terms of content validity, understandability, and practicality.

Keywords:

Monitoring; Tailoring; Diabetes; Weight Loss

Introduction

As a chronic metabolic disease, diabetes cases are increasing nationally and globally. There is no definite cure for diabetes and it requires continuous lifelong care [1]. Studies have shown that 85% of diabetics are overweight [2]. It is commonly suggested that obesity contributes to the accelerated ischemic vascular disease and mortality of patients with type 2 diabetes [3]. It has been proven that weight loss improves survival in obese type 2 diabetic patients and improves life expectancy [4]. Centers for Disease Control and Prevention, states "The good news is that no matter what your weight loss goal is, even a modest weight loss, such as 5 to 10% of your total body weight, is likely to produce health benefits, such as improvements in blood pressure, blood cholesterol, and blood sugars" [5]. According to the World Health Organization (WHO), diabetes and obesity are preventable chronic diseases [6]. A main requirement of chronic disease management is continuous follow-up and adherence to therapeutic plans. In order to assess the patient's compliance we need to monitor the degree to which a patient correctly follows medical advice. Since weight loss takes time and each doctor-patient visit requires recording several measurements, eventually recorded data increases in size, and subsequently its analysis becomes time consuming and arduous. Automatic data abstraction and providing it to the physician not only saves time and increases the physician's awareness about the patient's status, but also helps him/her to make informed decisions.

We have designed a system that interprets the patients' condition in each phase of the treatment course according to guidelines as well as experts' knowledge and experiences and have presented it as a textual report. Text generation is one of the key issues in NLP (Natural Language Processing) and follows either of two approaches: deep or shallow techniques. The former is based on semantic understanding and linguistic rules. The latter, however, uses simpler text generation methods that are domain-dependent [7]. Choice of the right approach depends on the purpose of the study. In shallow methods, there is a main template for the output text in which the pieces of data items are used to complete the statements based on the idiosyncratic characteristics and conditions of each patient. For each characteristic, a particular snippet has been incorporated within the system's knowledge-base that is inserted in the right place in the text according to the defined criteria. In the present research, according to the formal structure and specified format of messages in medical sciences, the template-based method was used.

Most technology-based weight loss interventions are designed with the consumer-centric approach. A systematic review by Levine et al. shows that technology cannot replace human interaction, but serves as a complementary and facilitator tool to improve patient-physician communication [8]. One of the most important features of the proposed system is the possibility of generating two separate reports, one for physicians, and another for the patients. For the physician, the report provides an overview of the patient's follow-up events and for the patient, the report consists of performance feedback and recommendations that are relevant to each specific patient. A body of research has indicated that personalizing medical advice in accordance with one's needs, demands, and level of understanding would increase its efficacy [9].

In this paper, we have reported on a system designed by the authors that generates personalized reports for the diabetic patients' adherence to weight loss recommendations. The method has been described, followed by the results of the analyses, and finally a discussion of the results and the probable causes and factors involved are presented.

Methods

System input is comprised of the data related to weight measurement and the related date along with the domain knowledge required to interpret it. System output is comprised of two separate written reports: one for the doctor and the other for the patient that is editable by the doctor. After the data analysis, the system gives an interpretation of the patient's progress during the course of monitoring (Fig. 1).

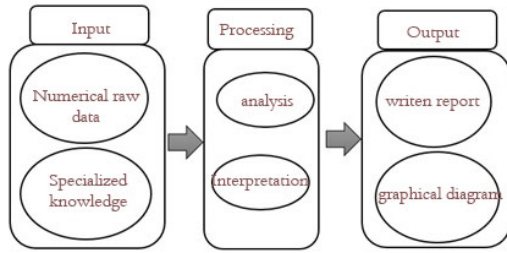


Figure 1- Structure of the weight loss interpretive system

The system has been designed in three phases: 1) specifying report structure, 2) rule-based data interpretation, and 3) generating textual/graphical reports for the patient and physician.

Phase I: Specifying Report Structure

The report addressed to the physician explores the patient's condition from four aspects:

- Patient's current status vs. expected status,
- Evaluation of the patient's compliance throughout time (e.g. progress in loss of weight and the extent of following the treatment program),
- Estimation of the approximate time to reach ideal weight (prediction), and
- Providing useful information about the impact of weight loss on general health.

The first step in designing the weight monitoring system is the identification of data items based on the physician's needs. These items were specified according to a review of literature [10] and interviews with domain experts and include: patient's ideal weight based on his/her demographic and clinical features (such as age, blood glucose, prescribed medication and so on), BMI, length of monitoring, frequency of weight records, percentage of falling behind or exceeding the schedule, and the ideal time for achieving the desired weight.

The other report has been generated to address the patient and includes personalized feedbacks and recommendations.

Phase II: Data analysis rules

A rule-based approach was used to interpret the data. Two credible knowledge sources were used to extract rules concerning weight loss in diabetic patients: expert panel and the guidelines' recommendations [10] [11]. Overall, 46 rules were found and represented as "if...then" statements in the system's knowledge base. Figure 2 shows sample rules for the BMI interpretation.

```

If BMI >= 16.5 AND BMI < 18.5 Then
Inter = "UnderWeight"
Elseif BMI >= 18.5 AND BMI < 25 Then
Inter = "Normal"
  
```

Figure 2- BMI interpretation rules

Rules are categorized in three groups based on their application:

- Rules concerned with patient's adherence to follow-up, like the measurement intervals and the standard amount of weight loss per session.
- Rules associated with the weight loss program, estimating the ideal weight, patient's status based on standard classification of obesity.

- Rules demonstrating the positive impacts of weight loss on diabetic patients' health.

Phase III: Report Generation

The structure and components of the model obtained from the previous phase were implemented in the Visual Studio.net framework and Microsoft Access database. In this system, each patient has a unique identification code. There is a data entry form that makes a record of the measurements made in each visit. A sample report output is shown in Appendix 1. As can be observed, the data have been presented in different graphical formats like tables, diagrams, and texts. The structure of the report has two parts. The above section is used by the physician and presents the information about the extent to which the treatment course has progressed by the patient. The section at the bottom of the page is addressed to patients and aimed to make them aware of one's progress. It was only availed to patients once confirmed by the physician. Figure 3 shows the text template that was completed with the appropriate patient's data.

In <@per1> of cases, the measurement intervals have been <@stat> which show the <@qual> rate of follow-up.

We expected a <@ideal> loss since the outset. As compared to the <@done> lost so far, there shows to be a <@rate> of <@status> the program.

Figure 3- Sample text template

Results

One way of assessing the textual content of an automatic generated report is to compare it to a human-made report. However, human beings act mentally to write a textual report. Therefore, it is not possible to develop a gold standard to compare the system results. The evaluation methods of automatic reports are divided into two groups:

- Intrinsic evaluation: the generated report is assessed in terms of the internal quality such as cohesion and communication of information.
- Extrinsic evaluation: impact analysis of the generated report within the context of diabetes weight management.

We followed a two-phase approach to evaluate the generated report, considering perspectives from the patients and doctors. To this aim, we developed and validated a 6-item questionnaire based on [12]. Three representative sample reports was also generated by the system, to demonstrate the most prevalent patterns of the diabetic patients' weight loss which consisted of irregular follow-up, follow-up as planned, and falling behind the program.

Table 1 – Physicians and Patients perspectives on inner quality of the generated report

Questions	Physicians	Patients
To what extend was the report clear and understandable?	86%	94%
To what extend did you find the information relevant?	92%	97%
To what extend did you find the text coherent and cohesive?	82%	88%
How adequate was the provided information?	75%	81%
To what extend did you perceive the report as practical?	82%	93%
To what extend was the information new to you?	77%	89%

In the first phase, the sample report was submitted to 12 nutrition specialists and eight diabetic patients along with the 5-point Likert scale questionnaire, ranging from “Strongly Disagree” on one end to “Strongly Agree” on the other with “Neither Agree nor Disagree” in the middle. Table 1 shows the questionnaire with mean scores obtained by each group of physicians and patients, respectively. Values in the table are in percentage obtained through $(\text{mean scores} \times 100) / 5$. As the primary aim of our system is to assist physicians dealing with large volumes of data, we also compared the physicians' awareness about the patient's visit/medical history in the current format and the generated report proposed in this study. To accomplish this, we made an information extraction template composed of essential data items needed for informed decision making, which was determined by physicians in a delphi study. The information extraction template consists of five data items including: patient's compliance status to visit sessions, patient's compliance status to the weight loss program, patient's weight loss trend status, patient's violation from ideal weight, and expected time to reach the ideal weight. We also determined a time limit for the physicians to fill out the template form. Physicians were first provided with the usual raw data of each visit session and asked to fill out the information extraction template within the pre-defined time limitation. They were then provided with the computer-generated report, and again asked to fill out the template form within the same time.

This assessment was done with 12 specialist for the three sample reports. The proportion of correct responses for each condition are shown in Table 2.

Table 2 – Physicians' correct response rate with current format vs. the system's generated report

Information Item	Current Format	Generated Report
Compliance to visit	53%	100%
Compliance to program	64%	100%
Weight loss trend	56%	94%
Difference from ideal weight	91%	100%
Time to reach the ideal weight	77%	100%

In the second phase, a semi-structured interview was conducted with both patients and doctors respectively and their opinions and perspectives regarding the quality and applicability of the report were assessed with open-ended questions. The doctors approved the usefulness of the content and agreed that it facilitated their decision-making. In this respect, they maintained: “Currently, to access the patient's information, disperse information should be looked for in the patient's records. It is both time-consuming and erroneous.” This report helps to avail the foremost information in the shortest possible time to the experts. Another evaluator suggested sectionalizing the content and highlight it in different colors in order to help find the intended information in shortest time. Participants also approved the adequacy of the content, but drew attention to the necessity of adding other factors related to weight loss such as the amount of calories intake, physical activity, percentage of fat and water in the body to practically apply the program in the doctor's office. The experts found the patient's report section as a useful and practical aspect. Here is a comment in this regard:

“The patient's report is clearly elaborated and has a motivational effect. For more efficacy, reports can be sent to patients as a text message or email so that it prevents them from forgetting their treatment and encourages them to follow their treatment.”

Overall, the experts had a positive attitude towards using the system and the combination of text, tables, figures, and graphs to convey the results.

Discussion

Electronic interventions concerning weight loss such as text messaging, mobile, and web-based applications and other computer-based software have been primarily designed for patients' self-monitoring. The output of these programs generally displays one's progress in graphical diagrams and delegates the interpretation to the reader. This study is the first in its kind that provides an automatic explanation about patient's status in text and graphic format both for the physician and patients.

Results show that the overall scores in the physician's perspective were lower than patients in all six scales, which can be due to their more expertized view. The lowest score for both groups of audiences belongs to information adequacy, which indicates some information that the users were concerned about was neglected in the study. Comparing the current practice and the automatic generated report revealed that the physicians outperformed in working with structured, interpreted information rather than seeking patterns in mounds of raw, unstructured data.

Monitoring medical data is an essential task for clinicians in domains such as the ICU data stream [13] or diabetic patients' blood glucose [14]. Numerous studies have documented the short-term benefits of weight reduction for patients with type II diabetes [15][16]. The weight loss monitoring report structure proposed in the present paper has been designed for use in a doctor's office and helps the doctor to interpret the patient's progress data precisely and in the least time.

Among the interpretive systems in medical sciences is PUFF [17]. It takes a rule-based method and interprets the results of pulmonary tests such as the lung volume test, patient's ability to inhale and exhale air, absorbing oxygen in blood and repelling the CO₂, and the severity of pulmonary diseases. It was practically used in the Pacific Medical Center in San Francisco and the evaluation results indicate an 89% correspondence between the system and doctor's diagnoses. The system designed in the present research acts as a complementary support in decision-making and provides a description of the measures that help the doctor to make more informed decisions. However, the system does not directly interfere in the physician's decision-making role. In fact, one advantage of the automatic report generation systems, which has made them more popular in medical practice as compared to CDSS (Clinical Decision Support System), is that it does not give direct medical advice [18]. Instead, it provides a description of events and gives more emphasis on important issues that reduces doctors' resistance using them [19].

In the AI systems categorization, we can consider our system as a data-to-text generator. BabyTalk is a data-to-text summarizer system using natural language generation techniques to produce automatically nursing end-of-shift reports. It was developed in 2012 by Hunter et al. in the neonatal's ICU of England's Empire Clinic [13]. To evaluate this system, the nurses in the ward rated the textual report generating system in three dimensions: comprehensibility, accuracy, and usefulness. The results showed 90% comprehensibility, 70% accuracy, and 59% usefulness. The evaluation approach in the present research resembled the investigation just mentioned. However, the qualitative approach did not allow a comparison of the results.

Summarizing large volumes of data and interpreting it in an easy understandable fashion can be considered as the strong point of this study, which enables doctors to identify important patterns and trends in the shortest possible time, leading to enhanced patient safety, improved doctor-patient relationship, and improved physician decision-making.

Our system generates two different reports organized for the patient and the physician, which can be considered as a unique contribution not seen in previous studies.

Since the present research is a pilot study, the limited number of participants and the limited exposure time to the system are considered as the limitation of this research, which can affect the generalizability of the findings.

Our suggestion for future studies is to conduct a need assessment investigation to determine and address users' priorities and wishes prior to system development. Due to the importance of psychological issues in the care of diabetic patients, we suggest considering this aspect in future researches. We also suggest investigating the long-term effect of the system on the patients and physicians' performance.

Conclusions

The present research has shown that the quality of the report made by the system directly improved the physician's performance and can potentially contribute to better medical outcomes. The criterion for evaluating the present system has been the qualitative content analysis of the report. The majority of the existing weight loss programs measure weight lost as an index of the efficiency of the intervention.

The results have suggested that automatic data abstraction is required by specialists in different healthcare domains, provides more information to the physician, and saves time for better patient-physician communication. It is recommended that researchers further investigate other relevant factors adding up to the generated report such as weight management, the amount of calories intake and physical activities. How each factor changes through time and how the other factors are correlated need to be investigated as well. The present findings have shown that the cooperation of healthcare specialists and IT researchers can lead to the design of practical systems that are warmly received by the medical staff. The present researchers believe that the proposed model can be applied in all monitoring programs of patients with chronic diseases as is required in the healthcare world today.

Acknowledgements

This study is the first author's PhD dissertation, which has been supported by a grant [Grant # 950392] from the Mashhad University of Medical Sciences Research Council. The authors would like to acknowledge the diabetes specialists from the Endocrine Research Center for their participation in the evaluation of the interviews.

References

- [1] J.E. Shaw, R.A. Sicree, and P.Z. Zimmet, Global estimates of the prevalence of diabetes for 2010 and 2030, *Diabetes research and clinical practice* **87** (2010), 4-14.
- [2] A. Powell, Obesity? Diabetes? We've Been Set Up, *Harvard Gazette*, [cited 2017 Apr 8], Available from: <http://news.harvard.edu/gazette/story/2012/03/the-big-setup/>
- [3] M. Lean, J. Powrie, A. Anderson, and P. Garthwaite, Obesity, weight loss and prognosis in type 2 diabetes, *Diabetic medicine* **7** (1990), 228-233.
- [4] R.R. Wing, R. Koeske, L.H. Epstein, M.P. Nowalk, W. Gooding, and D. Becker, Long-term effects of modest weight loss in type II diabetic patients, *Archives of internal medicine* **147** (1987), 1749-1753.
- [5] Losing Weight, Healthy Weight, [cited 2017 Apr 8], Available from: https://cdc.gov/healthyweight/losing_weight/index.html
- [6] World Health Organization. *Preventing chronic diseases: a vital investment*. World Health Organization, 2005.
- [7] S. Busemann, H. Horacek, A flexible shallow approach to text generation, *arXiv preprint* (1998).
- [8] D.M. Levine, S. Savarimuthu, A. Squires, J. Nicholson, and M. Jay, Technology-assisted weight loss interventions in primary care: a systematic review, *Journal of general internal medicine* **30** (2015), 107-117.
- [9] M. Tara, Aspects of Information Tailoring in the 21st Century, in *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour, Editor (2015), p. 4042-4052.
- [10] L.S. Wieland, L. Falzon, C.N. Sciamanna, K.J. Trudeau, S. Brodney, J.E. Schwartz, and K.W. Davidson, Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people, *Cochrane Database Syst Rev* **8** (2012).
- [11] R.R. Wing, W. Lang, T.A. Wadden, M. Safford, W.C. Knowler, A.G. Bertoni, J.O. Hill, F.L. Brancati, A. Peters, and L. Wagenknecht, Benefits of modest weight loss in improving cardiovascular risk factors in overweight and obese individuals with type 2 diabetes, *Diabetes care* **34** (2011), 1481-1486.
- [12] L. Hasan, E. Abuelrub, Assessing the quality of web sites, *Applied Computing and Informatics* **9** (2011), 11-29.
- [13] J. Hunter, Y. Freer, A. Gatt, E. Reiter, S. Sripada, and C. Sykes, Automatic generation of natural language nursing shift summaries in neonatal intensive care: BT-Nurse, *Artificial intelligence in medicine* **56** (2012), 157-172.
- [14] R. Pivovarov, D.J. Albers, G. Hripsak, J.L. Sepulveda, and N. Elhadad, Temporal trends of hemoglobin A1c testing, *Journal of the American Medical Informatics Association* **21** (2014), 1038-1044.
- [15] K. Fujioka, Benefits of moderate weight loss in patients with type 2 diabetes, *Diabetes, Obesity and Metabolism* **12** (2010), 186-94.
- [16] L. Van Gaal, A. Scheen, Weight management in type 2 diabetes: current and emerging approaches to treatment, *Diabetes Care* **38** (2015), 1161-72.
- [17] J.S. Aikins, J.C. Kunz, E.H. Shortliffe, and R.J. Fallat, PUFF: an expert system for interpretation of pulmonary function data, *Computers and biomedical research* **16** (1983), 199-208.
- [18] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes, Automatic generation of textual summaries from neonatal intensive care data, *Artificial Intelligence* **173** (2009), 789-816.
- [19] A. Gatt, F. Portet, E. Reiter, J. Hunter, S. Mahamood, W. Moncur, and S. Sripada, From data to text in the neonatal intensive care unit: Using NLG technology for decision support and information management, *AI Communications* **22** (2009), 153-186.

Address for correspondence

Mahmood Tara, MD, PhD. Address: Pardis Daneshgah, Park Square, Mashhad, 917-7948-564, Islamic Republic Of Iran. Phone: 98 5138002429 Email: taram@mums.ac.ir

Appendix 1: A report sample



WEIGHT LOSS REPORT



Patient's code: 10

Date: 12/19/2016

Patient's name: Ali Ahmadi	Age: 36 yrs	Type of weight loss program: Normal	Beginning Date: 12/19/2016
----------------------------	-------------	-------------------------------------	----------------------------

Doctor's report:

Ali Ahmadi's ideal weight is 74 kilograms. So, right now, he has 28 kilograms of extra weight. His BMI is 33 and belongs to class 1.

The frequency of recording his weight has been 11 since 18 months ago which shows a moderate rate of follow-up. Moreover, in 60% of cases, the measurement intervals have been delayed which show his moderate rate of follow-up.

Considering the standardized loss of 1.5 kilogram a week, ideally speaking, we expected a 27-kilogram loss since the outset. As compared to the 8 kilograms lost so far, there shows to be a 70% of falling behind the program.

Ideally, one is expected to reach the optimal weight within the forthcoming 19 weeks. However, due to the current trend, this can be expected to happen in 29 weeks.

The weight loss rate has been estimated in all cases to be below the threshold level which indicates that the patient fell behind the plan.

**Patient's Report:**

You have managed to reduce 30% of the amount you are overweight according to the time spent from the starting point, this value should be around 90%. By following this process it takes seven and a half months to reach the ideal weight. While if you do more regular weight measurements and follow the prescribed diet and weight loss it reduce to four and a half months. Here you have a weight loss of 8 kg, significantly reduced the risk of heart disease and have added at least 2 years to your life expectancy.

More Recommendations:

1.
2.
3.

Physician's Name

An Infrared Thermal Images Database and a New Technique for Thyroid Nodules Analysis

José R González^a, Charbel Damião^b, Aura Conci^a

^aComputing Institute, Federal Fluminense University, Niterói, Rio de Janeiro, Brazil,

^bAntonio Pedro University Hospital, Federal Fluminense University, Niterói, Rio de Janeiro, Brazil

Abstract

Thyroid nodules diseases are a common health problem and thyroidal cancer is becoming increasingly prevalent. They appear in the neck and bottom neck region, superficially over the trachea. Cancer tissues are characterized by higher temperatures than surrounding tissues. Thermography is a diagnostic tool increasingly used to detect cancer and abnormalities. Artificial intelligence is an approach which can be applied to thyroid nodules classification, but is necessary to have a proper number of cases with proven diagnosis. In this paper, a new database that contain infrared thermal images, clinical and physiological data is presented. The description of each nodule per patient, and the acquisition protocol (based on Dynamic Infrared Thermography approach) is considered as well. A semi-automatic method for image registration was implemented to pre-process the thermograms and a new method for the Region of Interest (ROI) extraction is proposed. Moreover, the obtained ROI results are confirmed by medical specialists and turned available for future comparison with other works.

Keywords:

Thyroid Nodule; Infrared Rays, Diagnosis; Thermography

Introduction

The thyroid is an endocrine gland located at the human neck, next to the thyroid cartilage and over the trachea. Thyroid nodules are a common clinical problem and differentiated thyroid cancer is becoming increasingly prevalent. Epidemiologic studies have shown the prevalence of palpable thyroid nodules to be approximately 5% in women and 1% in men living in iodine-sufficient parts of the world. In contrast, high-resolution ultrasound (US) can detect thyroid nodules in 19%–68% of randomly selected individuals, with higher frequencies in women and the elderly [1]. The clinical importance of thyroidal nodules investigation is for exclusion of thyroid cancer, which is prevalent in 7 to 15% of cases, depending of factors as age, sex, exposition to radiation, and familiar history [2].

The temperature distribution on the human skin can be measured by using infrared cameras and stored in a thermal image. These images represent a temperature pattern of the body, which is highly symmetric around the vertical axis of the sagittal plane. Variations of this symmetry in a serial imaging can constitute a sign of abnormality. Thermal imaging is an inexpensive and non-invasive technology that has been applied in many fields, including sport medicine, forensic medicine, anesthesiology, peripheral vascular diseases, and cancer and breast diseases diagnosis [3]. Heat pattern measurement by thermography, is fast, non-contact and non-invasive [4].

A malignant tumor is related to abnormal growth of cells, invading tissues and spreading to other regions of the body.

These tumors need nutrients to grow and lead to the development of new blood vessels around them (angiogenesis) [5;6]. Due to increased blood flow, tumors frequently present higher temperature than the surrounding region, which can be acquired by thermography and used to aid in the diagnosis of malignant thyroids tumors [7]. In addition, toxic autonomous nodules tend to be large (>2.5 cm) and this may increase the possibility of detection [10].

Thermography detects physiologic or functional changes on the surface of the skin and the Pennes's equation [8] or artificial intelligent methods can be used for detect the possible internal cause of such thermal pattern. Thermograms are sensitive to environmental changes in temperature, humidity and air circulation. Moreover, they should follow the same acquisition protocol to standardize the process and minimize the intra and inter patient thermograms variations. Additionally, the thermograms must be preprocessed in a proper way to minimize possible errors. Discriminated features need be extracted for neck region (Regions of Interest, ROI) and used as input in algorithms for pattern recognition. This work proposes a protocol for infrared image acquisition; a database for thyroid study with infrared images and clinical data and a method for ROI autonomous identification and image registration of thyroid region.

Background

The thyroid nodule diagnosis using thermography has been studied in few works in the literature. A numerical analysis of the area by Finite Element Method is proposed by Helmy et al. [3, 9]. The authors also compare numerical simulation with a thermal image of the thyroid gland for the same patients [10].

Gavriloua et al. [11;12] explain details of a system for infrared image acquisition to study thyroid nodules and analyze them by Penne's equation. However, few infrared images are explored to find infrared signatures that can be used as descriptors of the thyroid tumors [11]. They use the ABCDE investigation method (based on Asymmetry, Border, Color, Diameter and Evolution of the contour) and affirm that their method can correctly identify 89.3% of the investigated patients with thyroid cancer. Moreover, they apply fractal analysis to quantify the irregularity in size and shape of thermal signatures of tumors and use self-similarity and lacunarity features [13]. Same authors purpose an improved method for IR image filtering [14] as well. This study is aimed at developing a numerical scheme which significantly reduces the computer time for thermal image denoising with edge preservation. This filter allows physicians to assess faster than using other anisotropic diffusion filters, the contour shape, to locate the outbreaks in ROIs. Other filtering types are used for improve the thermals images and for Empirical Mode Decomposition [15].

The results of thermography of thyroid dysfunctions are presented as an alternative clinical diagnostic technique

considering 37 patients of Lagos University Hospital (Nigeria) with confirmed thyroid diseases and 16 volunteers as control [16]. Cytological methods were used for confirming the thyroid disease while the images of skin temperature maps were obtained using the FLIR Infrared Camera operated in self automatic calibrated mode. Thermograms were compared with ultrasound exams. The mean temperature \pm standard deviation of 36.63 ± 0.56 °C was obtained for hyperthyroid, 34.93 ± 0.32 °C for hypothyroid, 35.76 ± 0.49 °C for control groups respectively. The variations in the mean skin temperatures of malignant are 37.63 ± 0.29 °C and of benign 36.21 ± 0.73 °C nodes. The results show that thermography could be relevant and if used in combination with other imaging modality, could play a vital role in the diagnosis of thyroid diseases. This study also confirmed that females are more susceptible to thyroid diseases than males and that hyperthyroidism is more frequent than hypothyroidism.

In [17], an automatic methodology to classify some thyroid disorders as hypothyroid or hyperthyroid is proposed. First the infrared images of neck of thyroid patients are taken by using a thermal camera, then a median filter is used, and the images are enhanced by equalization. Segmentation of the images and ROI extraction are done by using modified Otsu Technique. Texture features were extracted by using Gabor's filters and wavelet. Contrast, Homogeneity, Entropy and Energy features from the gray level co-occurrence matrix are used as well. Classification is done by a Multilayer Perceptron Network (MLPN).

Other system to detect the thyroid gland disease using thermograph is proposed [18:19], where the patients' neck are captured by using a FLIR-E30 camera. These images were filtered by using median filter, and enhanced by histogram equalization. The segmentation of the images is done using Otsu's Technique, features are then extracted from region of interest: mean, variance of the low-low-frequency sub-band using the Haar Wavelet transform, Coefficient of Local Variation, difference between the number of pixels in a block and the ratio of the sum of pixel intensities in the block to the maximum in the block, Normalized Multi-Scale Intensity Difference (defined as the differences between the pixel pairs with horizontal, vertical, diagonal, and asymmetric-diagonal directions), and the area occupied by the region of interest. Finally, the thyroid images are classified as hypothyroid, hyperthyroid or normal using a Bayesian Classifier [19].

Methods

Recent applications of thermography in medicine showed the effectiveness of Dynamic Infrared Thermography (DIT) in medical research. The DIT is a method for monitoring the dynamic response of the skin temperature after thermal stress in a period of time [20].

DIT Protocol for thyroids thermograms

We recommended that at least thirty minutes before the examination, the patient should avoid: alcohol, caffeine, physical exercise, nicotine, and should not apply any cream, oil or chemical substance to the neck region. Initially, the patient was asked to sit still in order to minimize the possible displacements; the distance from the camera to the patient was 0.5 to 0.6 meters. The relative humidity of air and room temperature were recorded and inserted as parameters in the camera. Room temperature was maintained between 22 °C and 25 °C, no doors and windows openings and with only fluorescent bulbs. Patients were requested to remove earrings, necklaces or any other accessory that can be seen in the thermal image and use a hair band. Body temperature was checked using a thermometer. Patients were seen in the room twenty minutes before starting the examination and were seated with

their head tilted slightly back and looking up while capturing images. An air flow (electric fan) was directed to the patient, when the mean temperature of the skin decreases to 29 °C, and this ventilation was stopped and the sequential acquisition was started. The cooling of the neck region improved the thermal contrast between healthy and unhealthy tissues. One image was captured every 15 seconds over five minutes, producing one sequence of twenty images.

Database

The database presents data and images from volunteers and patients of Federal Fluminense University Hospital (named Antonio Pedro University Hospital – HUAP - www.uff.br/huap/) in Niterói, Rio de Janeiro, Brazil. In this work, we use data from 25 volunteers (twenty with benign thyroid nodules, and 5 healthy as control group). The average age was 54 and female was the predominant gender. A FLIR thermal camera model SC620 was used to capture the thermograms. The sensitivity of the camera was smaller than 0.04°C and images had a dimension of 640x480 pixels. The image acquisition and their use for research was approved by the Ethical Committee of the HUAP and registered at the Brazilian Ministry of Health under number CAEE: 57078516.8.0000.5243. The database is accessible at <http://visual.ic.uff.br/thyroid/>.

Processing methods for patients classification

The proposed methodology for this is divided into four stages. First, the thermograms are registered to minimize the thermograms acquisition errors. Then, a neck and sub-neck region where thyroid nodules are located is segmented (the region of interest is segmented). Globals and locals features from ROI of each image are extracted to create few time series by patient. In the last stage, decision is performed by pattern classification of time series.

Image Registration

The protocol for image capture was designed to standardize the process and minimize the errors that can occur during the thermograms capturing process, but movements of patients can occur, causing misalignments between the images (Figure 1), and error in further processing.

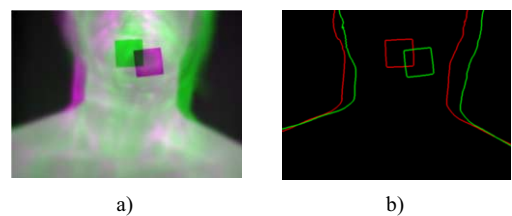


Figure 1 - Movements of a patient during the thermograms acquisition.

Two general types of movements can be identified: i) full-body modification, that include lateral movements to the left or right, to the top or bottom, or combinations of both; and ii) local modifications, that include movements related to perspective and distortion, such as tilts of the head to the back or front.

We use rigid transformation (translation and rotation) to correct the full-body modification. The first image is considered as reference or source, and each one of others as destination. At least two correspondent points are required to determinate the rigid transformation T (Equation 1) that maps the coordinates of the pixels $x = (x_1, x_2, 1)$ of source image into a new coordinates $x' = (x_1', x_2', 1)$ of sensible image, where $x' = T \cdot x$.

$$T = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

The correspondence of points can be determined by algorithms or manually. In this work, the correspondence points are selected manually.

ROI extraction

After the registration process we extract the region of interest to focus the analysis in the neck region (thyroids). The ROI of the reference image (the first) is extracted using the algorithm proposed in this section. ROI is then identified for each of the other registered images.

The automatic algorithm proposed to ROI extraction is based on locating the neck region and the bottom neck area. The neck region is assumed be in the narrowest part of the patient body. In the algorithm, the left upper corner of image is considered the origin of coordinates and it is composed by six stages, as follow:

Algorithm

1. The image is transformed to black and white using the Otsu method [21].
2. Background points are removed. Pixels belonging to patient area are replaced by the temperature of the original image.
3. Determine the left L and right R edges of patient using left-to-right and right-to-left scanning and gray level information of pixels in the black and white image.

The laterals limits of neck region are formed by the pixels $x \in L$ and $x' \in R$, with $x = (x_1, x_2)$ and $x' = (x'_1, x'_2)$ such that:

$$\begin{aligned} (ANW - 0.1ANW) &\leq abs(x_1 - x'_1) \\ &\leq (ANW + 0.1ANW) \end{aligned} \quad (2)$$

and $x_2 = x'_2$.

The ANW values is the neck average width of patient in the image, and is calculated from the $N=80$ minimal differences between all pixels $x \in L$ and $x' \in R$, with $x_2 = x'_2$:

$$ANW = \frac{1}{N} \sum abs(x_1 - x'_1) \quad (3)$$

4. The upper limit of ROI is formed by the pixel row that link the start of left neck edge with the right neck edge.
5. The left (right) limit of bottom neck area is formed by the pixels in the same image column that the last one pixel in the left (right) neck boundary determinates in step 5. The bottom limit of the bottom neck area is formed by the pixels in the last row of the image and into the left and right limits of bottom neck area.

In Figure 2 are shown in red line the ROIs result from two different patients.

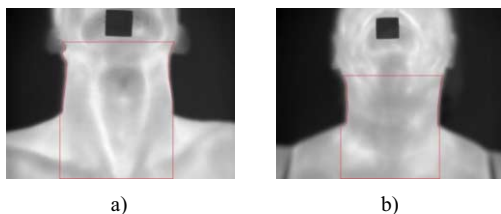


Figure 2 - Automatic ROI segmentation.

Features extraction

Thermograms can be processed as time series of temperatures if errors during the acquisition process are minimized. A thermogram is a two-dimensional map of temperature, and can be seen as a gray level image if scaled correctly. Moreover, methods of image analysis are applied to process the thermograms and descriptor features extraction, and artificial intelligence methods for classification. Features from ROI are necessary for the artificial intelligence methods of nodules diagnosis. These features must differentiate anomalous and normal patterns on unhealthy and healthy areas. Thermals patterns in the healthy human body are highly symmetric regarding the vertical axis of the sagittal plane. To this end, the coefficient of asymmetric level is determined from all pixels of the ROI as below:

$$\begin{aligned} Assim(I) = & \frac{1}{mn} \left(\sum_{i=1}^m \sum_{j=1}^n \frac{\min_{p,q=-1:1} |I(i,j) - I(i+p,n-j+q)|}{255} + \right. \\ & \left. \sum_{i=1}^m \sum_{j=\frac{n}{2}+1}^n \frac{\min_{p,q=-1:1} |I(i,j) - I(i+p,n-j+q)|}{255} \right) \end{aligned} \quad (4)$$

where m and n are the number of row and column of the ROI, and $I(i,j)$ is the gray level intensity value of pixel in the position i,j in the image. This feature is based on the similarity of gray level intensity value of each pixel in one vertical half of ROI and its most similar in the respective 3x3 window in the other vertical half.

It was observed that areas with nodules are characterized by high and constant temperatures in the recuperation phase to the thermal stress applied on the patient. On the other hand, temperatures in healthy areas were gradually increasing. Based on this statement, three features were selected from the ROI of each thermograms of the patient, as below: after binary thresholding of ROI using an empirical threshold T the hotter areas of the ROI were identified. These features are: Normalized Mean Intensity, Standard Deviation of Intensities and Normalized Maximum Intensity. Such features are considered as a time series of size 20. Then, each patient was represented as a pattern formed by four time series.

Classification

We employed the k-Nearest Neighbors (k-NN) [22] algorithm to perform a classification using the time series form extracted features. The quantity of neighbors k used was $k=1$. The Euclidean distance among time series was used to determine the similarity between the thermograms of two patients. The classification of a patient is performed to identify the most similar object, i.e. the patient described by the closest time series.

Results

A database was created for thyroid nodules research. This contains 25 cases with proven diagnosis. All cases were collected under the proposed protocol.

The pre-processing step of thermograms is very important to avoid further errors. In this case, an algorithm based on rigid transformation to images registration is implemented and evaluated. Ninety-one percent (91%) of images were correctly registered, but it is important to consider evaluate non-rigid methods as well, because the method can fail in images with locals deformations.

A new method for neck and sub-neck ROI image segmentation was developed. All images and used data can be found in

<http://visual.ic.uff.br/thyroid/>. These results are confirmed by physicians. ROI were extracted correctly in the majority of used images.

Conclusions

This work performs an overall analysis of thermography images, focusing on thyroid infrared image acquisition, processing and analysis, with is a new field of study. We analyze data critically in order to extract relevant information that can help in artificial intelligence conclusion related to cancer diagnosis by using such images.

A new database was created and published for researches on thyroid nodules diagnosis. We proposed an autonomous ROI identification method, which is based on very simple fundamentals of computers vision.

As continuation of this work we will consider: (1) developing a automatic algorithm for selecting reference points (based on anatomical identification of the patients' neck to been used as correspondent points in image transformations in the registration process); (2) analyzing non-rigid methods for the image registration process; (3) include others features to be used in the classification and (4) promoting a classification step by using data mining algorithm.

Acknowledgements

The authors are grateful to the partial support of the Brazilians agencies: FAPERJ (E26/202.959/2016), CAPES and CNPq (Project: No. 201542/2015-2 PQ – CA: EM).

References

- [1] American Thyroid Association, 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer, *Thyroid*, Number 1, Volume 26, 2016. DOI: 10.1089/thy.2015.0020
- [2] Hegedüs, L., The thyroid nodule, *New England Journal of Medicine*, (2004), 1764-1771. DOI: 10.1056/NEJMc031436
- [3] A. W. Helmy, M. Rizkalla, M. Holdmann and P. Salama, Finite Element Analysis for Simulating a Hot Thyroid Nodule, *Circuits and Systems*, Proceedings of the 43rd IEEE Midwest Symposium, (2000), 1064-1067. DOI: 10.1109/MWSCAS.2000.951399
- [4] O. Faust, U. R. Acharya, E.Y.K. Ng, T. J. Hong, and W. Yu, "Application of infrared thermography in computer aided diagnosis." *Infrared Physics & Technology* 66 (2014): 160-175.
- [5] A. A. Lars and A. Virginia, Increased angiogenesis in papillary thyroid carcinoma but lack of prognostic importance, *Human Pathology* (2000), 439-442. DOI:10.1053/j-ip.2000.6548
- [6] M. Alves, J. Andrade, J. Cherri, T. Moriya, C.E. Piccinato and N. Okano, Papel da termografia na seleção de nódulos tireoidianos de indicação cirúrgica. *Arquivo Brasileiro de Endocrinologia e Metabolismo* (1988), 97-99.
- [7] A. W. Helmy, M. Holdmann, M. E. Rizkalla, and P. Salama, A novel approach for a non-invasive diagnostic technique for thyroid glands using thermographic system, *Circuits and Systems* 2000, Proceedings of the 43rd IEEE Midwest Symposium (2000), 1094-1097. DOI: 10.1109/MWSCAS.2000.951406
- [8] H. Pennes, Analysis of tissue and arterial blood temperatures in the resting human forearm, *Journal of Applied Physiology*, (1948), 93-122.
- [9] A. W. Helmy, M. Rizkalla, M. Holdmann and P. Salama, Finite element analysis for simulating a hot thyroid nodule. *Circuits and Systems* 2000, Proceedings of the 43rd IEEE Midwest Symposium, (2000), 1064-1067. DOI: 10.1109/MWSCAS.2000.951399.
- [10] A. W. Helmy, M. Holdmann, and M. Rizkalla, Application of thermography for non-invasive diagnosis of thyroid gland disease, *IEEE Transactions on Biomedical Engineering*, (2008), 1168-1175. DOI: 10.1109/TBME.2008.915 731
- [11] J. Rizkalla, W. Tilbury, A. W. Helmy, V. K. Suryadevara, M. Rizkalla, and M. Holdmann, Computer simulation/practical models for human thyroid thermographic imaging, *Journal of Biomedical Science and Engineering*, (2015), 246-256. DOI: <http://dx.doi.org/10.4236/jbise.2015.84024>

- [12] G. Gavriloiia, A. M. Ghemigian, and M. R. Gavriloiia, Infrared signature analysis of the thyroid tumors, *European Conferences on Biomedical Optics, International Society for Optics and Photonics*, (2009), 1-6. DOI:10.1117/12.831756.
- [13] G. Gavriloiia, A. Hurduc, A.M. Ghemigian, and R. Fumarel, Spatial-temperature high resolutionmap for early cancer diagnosis, *Multimodal Biomedical Imaging*, 1-8. DOI: 10.1117/12.809185
- [14] G. Gavriloiia, M.R. Gavriloiia, E. Sofron, and A. M. Ghemigian, Using Fractal Analyze of Thermal Signatures for Thyroid Disease Evaluation, *Proceedings SPIE, Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies*, (2010), 1-6. DOI:10.1117/12.882294.
- [15] B. M. Gavriloiia, C. R. Vizireanu, O. Fratu, C. Mara, D. N. Vizireanu, R. Preda, and G. Gavriloiia, Thermal image filtering by bi-dimensional empirical mode decomposition, *Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies*, (2015), 1-6 DOI:10.1117/12.2070375.
- [16] M. A. Aweda, A. O. Adeyomoye and G. A. Abe, Thermographic analysis of thyroid diseases at the Lagos University Teaching Hospital, *Nigeria, Advances in Applied Science Research* (2012), 2027-2032. <http://www.pelagiarresearchlibrary.com>
- [17] Santiago, V., "Diagnosis of Hypo and Hyperthyroid Using MLPN Network", *International Journal of Innovative Research in Science, Engineering and Technology* (2014) 14314-14323.
- [18] Mahajan, P., and Madhe, S., "Hypo and hyperthyroid disorder detection from thermal images using Bayesian Classifier", *Advances in Communication and Computing Technologies (ICACACT)*, 2014 International Conference, IEEE, pp. 1-4, 2014. DOI: 10.1109/EIC.2015.7230721
- [19] Mahajan, P., and Madhe, S., "Morphological Feature Extraction of Thermal Images for Thyroid Detection", *International Journal of Electronics Communication and Computer Engineering*, Volume 5, Issue 4, pp. 11-14, 2014.
- [20] L. F. Silva, A. A. Santos, R. S. Bravo, A. C. Silva, D. C. Muchaluat-Saade, A. Conci, Hybrid analysis for indicating patients with breast cancer using temperature time series, *Computer methods and programs in biomedicine* (2016), 142-153. DOI: <http://dx.doi.org/10.1016/j.cmpb.2016.03.002>
- [21] N. Otsu, A threshold selection method from gray-level histograms, *Automatic* (1975), 62-66.

Address for correspondence:

José Ramón González Montero
 jgonzalez@ic.uff.br; josergm86@gmail.com
 +55 21965913089

Automatically Identifying Topics of Consumer Health Questions in Chinese

Haihong Guo, Xu Na, Jiao Li

Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China

Abstract

In health question answering (QA) system development, question topic identification is crucial to understand users' information needs and further facilitate answer extraction. This paper presented a machine-learning method to automatically identify topics of health related questions in Chinese asked by the general public. We collected 2000 questions from Chinese consumer health website, and characterized them using 17 types of features such as lexical, grammatical, statistical, and semantic features. This method were applied to identify 6 health question topics of Condition Management, Healthy Lifestyle, Diagnosis, Health Provider Choosing, Treatment, and Epidemiology. The results showed the average F₁-scores of the above 6 topic identification were 99.63%, 99.13%, 98.55%, 96.35%, 76.02%, and 71.77%, respectively.

Keywords:

Information Storage and Retrieval; Machine Learning; Medical Informatics.

Introduction

Generally speaking, a question answering (QA) system is composed of three key modules: question analysis, information retrieval, and answer extraction. Nowadays, QA systems widely used the natural language processing (NLP) techniques from question identification to answer generation [1]. As the first step, question classification plays a critical role in identifying users' information needs from the various and free expressions using natural language. Correct question classification would limit the retrieval scope and improve answer accuracy [2].

Automatic question classification methods are divided into rule-based and machine-learning based ones. Rule-based methods use the question format and domain knowledge as classification rules, which work well for classifying "5W1H" (i.e. what, where, when, who, why, how) questions but failed to solve complicated ones [2-3]. The machine-learning based methods are effective in learning classification rules from large-scale real corpus and delivering more objective and reliable results. The features of machine-learning based question classification include the following 5 categories [2,4]: a. lexical features such as bag-of-words (BOW) and part-of-speech (POS); b. grammatical features such as interrogative words (IW), noun/verb phrases and their grammatical functions; c. semantic features such as controlled terminologies and their semantic types (ST); d. pragmatic features such as emotional words and context scenarios; e. statistical features such as question length (QL), word length (WL), and term frequency-inverse document frequency (TF-IDF) etc. The widely used algorithms include support vector machines (SVMs), naive bayesian, decision tree, maximum entropy, logistic regression, and conditional random fields.

Several studies have been conducted for question classification in health and medical field in English, such as identifying general topics of clinical questions [5], distinguishing answerable and unanswerable intensive care unit (ICU) questions [6]; separating consumer health questions from professional medical questions [7], and classifying question types for consumer health questions [8]. Chinese question processing has its challenges in word segmentation, feature representation and selection, and classification algorithm improvement[3]. Some efforts have been made towards the general Chinese question classification, for instance, developing a question property kernel function which combines syntactic dependency relationship and POS [9], designing an alternating and iterative One-against-One algorithm [10], and developing a semi-supervised method with jointly learning question and answer representations [11]. It is worth noting that questions of health and medicine have their own characters such as circumscribed content, unique terminologies, and complex expression [12]. Few works have been investigated on Chinese consumer health question classification.

In this study, we designed a machine-learning method to automatically identify topics of consumer health questions in Chinese.

Materials and Methods

Data Set

A total of 2000 questions were manually annotated with one or more categories by the classification schema of consumer health question built in [13], and these questions were randomly selected from nearly 100 thousand hypertension-related messages posted by consumers on a Chinese health website (xywy.com). The distribution of questions on the primary level categories is shown in Table 1.

Table 1 - Topics and Distribution of the Questions

No.	General Topics	Positive	Negative	Total
1	Diagnosis	600	1400	2000
2	Treatment	1167	833	2000
3	Condition management	136	1864	2000
4	Epidemiology	233	1767	2000
5	Healthy lifestyle	278	1722	2000
6	Health provider choosing	45	1955	2000
	Total	2000	2000	2000

Feature Selection

We explored various features for machine-learning, including lexical, grammatical, semantic, and statistical information obtained from the initial BOW.

- BOW: Word segmentation was executed by Rwordseg [14], a Chinese word segmentation tool under R environment based on the Institute of Computing Technology, Chinese Lexical Analysis System (ICTCLAS), which reached a precision of 97.58% on the 973 experts testing [15].
- POS: The Chinese POS tagging was executed by the Stanford Parser (Version 3.3.1) [16].
- IW, noun head chunks (NHC), verb head chunks (VHC), noun rear chunks (NRC), verb rear chunks (VRC), IW + NHC/VHC, and NRC/VRC + IW: A dictionary of 42 Chinese interrogative words was developed manually based on baike.baidu [17] and general types of Chinese consumer questions that summarized by our former research [13]. NHC/VHC represents the first noun/verb after IW in a question. While NRC/VRC represents the last noun/verb before IW. They are likely to be the dependent word of the IW that help to express the semantic information of the question [18].
- Chinese Medical Subject Headings (CMeSH) concepts and ST: Medical concepts in the question and their ST were recognized by using the controlled vocabulary of CMeSH [19]. Three ST were considered: *Disease*, *Drug*, and *Symptom*.
- Keywords: Three ways to extract the keywords of a question were explored in this study: a. taking the first k words with maximum TF; b. taking the first k words with maximum IDF; c. taking the first k words with maximum TF-IDF. The heuristic equation (1) developed by Cao YG et al [5] was adopted to calculate k , which was based on their observation that the number of keywords increases when the question length increases.

$$k_{q_i} = \begin{cases} \frac{n_{q_i}}{6} + 1 \\ 3, n_{q_i} > 12 \end{cases} \quad (1)$$

Where k_{q_i} is the number of keywords extracted from the i^{th} question, and n_{q_i} is the number of words in the i^{th} question.

- Statistical features: including QL, max WL, min WL, average WL, max TF, min TF, average TF, max IDF, min IDF, average IDF, max TF-IDF, min TF-IDF, and average TF-IDF. The corpus used to calculate the IDF of each word contained nearly 100 thousand hypertension-related messages collected by our former research [13].

This study aimed to select the most discriminative features from the enormous amount of dimensions of feature space, some of which may either do nothing or degrade the performance of the classifiers. Φ -score that measures the discrimination of two sets of real numbers is a convenient technique used to achieve effective feature selection, which is defined as following [20].

Given training vectors $x_s, s = 1, \dots, m$, if the number of positive and negative instances are n^+ and n^- respectively, then the Φ -score of the t^{th} feature is defined as equation (2).

$$(t) = \frac{(\bar{x}_t^{(+)} - \bar{x}_t)^2 + (\bar{x}_t^{(-)} - \bar{x}_t)^2}{\frac{1}{n^+ - 1} \sum_{s=1}^{n^+} (x_{s,t}^{(+)} - \bar{x}_t^{(+)})^2 + \frac{1}{n^- - 1} \sum_{s=1}^{n^-} (x_{s,t}^{(-)} - \bar{x}_t^{(-)})^2}; \quad (2)$$

Where \bar{x}_t , $\bar{x}_t^{(+)}$, $\bar{x}_t^{(-)}$ are the average of the t^{th} feature of the whole, positive, and negative data sets respectively; $x_{s,t}^{(+)}$ is the

t^{th} feature of the s^{th} positive instance, and $x_{s,t}^{(-)}$ is the t^{th} feature of the s^{th} negative instance. Numerator in the equation indicates the discrimination between the positive and negative sets, meanwhile, denominator indicates the one within each of the two sets. The larger the $\Phi(t)$ is, the more likely this feature is more discriminative. Therefore, we used $\Phi(t)$ as the feature selection criterion, and the implementation steps were as following:

1. Calculate $\Phi(t)$ for each feature;
2. Calculate the $\bar{\Phi}$ (average Φ) of each type of features, and further set it as the threshold of the corresponding feature type. We choose $\bar{\Phi}$ as the feature selection threshold because the distribution of Φ differ greatly between different types of features (Table 2), while this method can help to keep all the useful features in different types [21].
3. For each type of machine-learning features, select features with $\Phi \geq \bar{\Phi}$ within their affiliated type.

Classifiers

Since the questions were multi-labelled, the task in this paper was a multi-label classification problem, which was usually transformed into one or more single-label classification or regression problems [22]. We therefore transformed the task into six binary classification problems (one-versus-rest for each) so as to suit the SVMs [23], which was commonly used and claimed to be the best in related works [5-8]. We used machine-learning algorithms within the R project for statistical computing (version 3.3.1) [24] for automatic question classification, including SVMs, naive bayesian, decision tree, maximum entropy, logistic regression, and conditional random fields.

Training and Testing

As shown in Table 1, the distribution of consumer questions to different topics is skewed, with a large majority of questions assigned to the topic of *Treatment* (1167, 58.35%), and *Diagnosis* (600, 30%), while only 45 (2.25%) questions were assigned to the topic of *Health Provider Choosing*. To compare the performance of different binary classifiers, this study applied the under-sampling method for the majority class, which could ensure each classifier was trained and tested on the same number of “positive” and “negative” questions. For example, when we trained a binary classifier for *Diagnosis*, there were 600 questions as positive cases (Table 1), then we have to select 600 negative cases randomly from the total amount of 1400 negative cases to keep the balance. Therefore, the sample data was composed of two distinct cases: the positive and the negative.

We then split the sample data for each binary classifier into two sets, 30% for independent test, 70% for 10-fold cross-validation, in which the sample data was equally divided into 10-folds, one of them was used as testing data, meanwhile, the remaining 9 folds as training data. The cross-validation process was repeated 10 times (equal to the folds) and the average value and standard deviation was reported. The advantage of this method was that all cases in the sample data was used for both training and validation, and each case was used for validation exactly once [25].

Evaluation Metrics

The performance were measured by precision (P), recall (R) and F1-score ($F_1 = (2 \times P \times R) / (P + R)$), which were commonly used in text retrieval tasks.

Results

Machine-Learning Features Selection

We calculated the Φ -score of every feature for each binary classification, and found that their distribution between different types of features differed greatly. The performance of features with $\Phi \geq \bar{\Phi}$ was not less than that of all the features in the corresponding types, and some of them even higher than the latter. For example, the topic of *Diagnosis* (Table 2), the $\bar{\Phi}$ of bag-of-words was 0.0009 with a standard deviation of 0.0029, while the value of IW was 0.0021 and 0.0079 respectively. when took all the 6154 features in POS as input features, the average F_1 -score was 78.92%, while just took the 2140 features with $\Phi \geq \bar{\Phi}$, the performance increased to 79.49%. Similar cases can be seen in feature types of IW, and keywords with maximum TF, IDF and TF-IDF

respectively, etc. Which interestingly indicated that some of the features in each type either do nothing to the classifiers or have some side effects on it. It's quite important to reject these features because it could not only save the computing resources so as to increase the efficiency, but also improve the performance of the classifiers. Therefore, we set $\bar{\Phi}$ of each feature type as its feature filtering threshold, and selected the features with $\Phi \geq \bar{\Phi}$ as input machine-learning features, so as to keep all the useful features in different types and improve the performance of the classifiers. Thus, each classifier received a different feature set, and the number of features within them were shown in the third column in Table 4. For example, the MaxTF was an effective feature for the classifiers for *Treatment*, *Condition Management*, *Provider choosing*, and *Lifestyle*, but not for the classifier for *Diagnosis*.

Table 2 – Φ Distribution and Performance of the Classifier for Diagnosis with Each Feature Type

Levels	Feature Types	$\bar{\Phi}$	$\sigma(\Phi)$	n_{AF}	$n_{(\Phi \geq \bar{\Phi})}$	F_{1AF}	$F_{1(\Phi \geq \bar{\Phi})}$
Lexical	BOW	0.0010	0.0031	4967	1761	0.7809	0.6483
	POS	0.0009	0.0029	6154	2140	0.7892	0.7949
Grammatical	IW	0.0021	0.0079	97	17	0.6129	0.6545
	NHC	0.0010	0.0025	48	5	0.1578	0.0759
	VHC	0.0009	0.0014	19	4	0.0823	0.0703
	NRC	0.0008	0.0009	73	34	0.2774	0.2606
	VRC	0.0009	0.0008	22	12	0.0661	0.0515
	IW + NHC	0.0011	0.0029	328	36	0.5059	0.3539
	IW + VHC	0.0007	0.0007	312	133	0.4341	0.4324
	NRC + IW	0.0008	0.0012	315	119	0.4956	0.5132
	VRC + IW	0.0010	0.0020	318	48	0.5139	0.4901
	CMesh concepts	0.0010	0.0026	43	7	0.6284	0.6656
Semantic	CMesh ST	0.0063	0.0075	3	1	0.6229	0.0150
	Keywords (TF)	0.0006	0.0003	1510	551	0.1923	0.7648
Lexical & Statistical	Keywords (IDF)	0.0006	0.0003	1137	371	0.1983	0.5527
	Keywords (TF-IDF)	0.0006	0.0003	1208	389	0.1522	0.5970
Statistical	Statistical features	0.0070	0.0065	13	7	0.6242	0.5513
Total with duplicates replaced				15349	5311	0.7984	0.9855

Note: For each type of features, $\bar{\Phi}$ is the average Φ , $\sigma(\Phi)$ is the standard deviation of Φ , n_{AF} is the total number of features, $n_{(\Phi \geq \bar{\Phi})}$ is the number of features with $\Phi \geq \bar{\Phi}$, F_{1AF} is the average F_1 -score of the classifier using all the features within this type, and $F_{1(\Phi \geq \bar{\Phi})}$ is average F_1 -score of the classifier using only the features with $\Phi \geq \bar{\Phi}$ within this type.

Performance of General Topics Classification

The single reported result in this study was from SVMs in the kernlab package which worked best from all different classification algorithms available in the R project for statistical computing. The performance of topics classification of each classifier with all the 15,349 features in each feature type as shown in Table 3, and the performance with features which $\Phi \geq \bar{\Phi}$ of each feature type was shown in Table 4. The research findings showed that by dropping features under the threshold, we could not only removed 65.40% to 85.13% of the features, but also improved the performance of each classifier. For example, the average F_1 -score of the classifier for *Diagnosis* dramatically increased from 79.84% to 98.55%. The results of 10-fold cross validation on each binary

classifier affirmed the hypothesis we proposed in the former section.

As shown in Table 4 and Table 5, we reached excellent performances on the classification of most topics of consumer health questions in Chinese, including *Diagnosis*, *Condition Management*, *Healthy Lifestyle*, and *Health Provider Choosing*, with all the evaluation metrics (average precision, recall and F_1 -score) above 90%. We reached moderate performances on the classification of *Treatment* and *Epidemiology*, with most of the evaluation metrics between 70% and 80%. The standard deviation of each evaluation metric indicated that the performance of each binary classifier was relatively robust.

Table 3 – Performance of General Topics Classification of Chinese Consumer Health Questions with All the 15349 Features

General Topics	AvgPrecision	AvgRecall	Avg F_1	$\sigma(\text{Precision})$	$\sigma(\text{Recall})$	$\sigma(F_1)$
Diagnosis	0.8059	0.7917	0.7984	0.0330	0.0275	0.0249
Treatment	0.7922	0.7286	0.7582	0.0404	0.0365	0.0277
Condition management	0.7981	0.6896	0.7328	0.1039	0.1388	0.1007
Epidemiology	0.6866	0.7138	0.6978	0.0490	0.0938	0.0617
Healthy lifestyle	0.9004	0.6333	0.7414	0.0582	0.0696	0.0545
Health provider choosing	0.7633	0.7900	0.7629	0.1737	0.2183	0.1791

Table 4 – Performance of the Classifiers for Chinese Consumer Health Questions with Selected Features by 10-fold cross-validation

General Topics	N(Selected Features)	AvgPrecision	AvgRecall	AvgF ₁	σ (Precision)	σ (Recall)	σ (F ₁)
Diagnosis	5311	0.9718	1.0000	0.9855	0.0316	0.0000	0.0164
Treatment	4216	0.8111	0.7178	0.7602	0.0437	0.0648	0.0482
Condition management	3150	0.9929	1.0000	0.9963	0.0226	0.0000	0.0117
Epidemiology	4194	0.7421	0.7078	0.7177	0.0431	0.1360	0.0798
Healthy lifestyle	3656	0.9869	0.9964	0.9913	0.0313	0.0113	0.0166
Health provider choosing	2282	0.9800	0.9550	0.9635	0.0632	0.0956	0.0594

Table 5 – Performance of the Classifiers on the independent test sets

General Topics	Precision	Recall	F ₁
Diagnosis	0.9790	1.0000	0.9893
Treatment	0.7998	0.7348	0.7654
Condition management	0.9882	1.0000	0.9940
Epidemiology	0.7562	0.7000	0.7254
Healthy lifestyle	1.0000	0.9976	0.9988
Health provider choosing	0.9289	0.9714	0.9487

Discussion

Comparison with Related Works

This study explored the possibility of using classifiers to identify topics of Chinese consumer health questions automatically and the feature selection by machine-learning to improve the efficiency and performance of each classifier. Compared to other related studies on question classification (Table 6), there were three specialties in this study. Firstly, Chinese consumer health questions were chosen as the research object.

Secondly, some feature types explored were more abundant than others. Our work adopted almost all the effective features in the prior works with the UMLS concepts and ST replaced by CMeSH concepts and ST. We also explored three ways to extract the keywords of a question as the machine-learning features, i.e. took the first k words with maximum TF/IDF/TF-IDF as keywords. In addition, we added NRC/VRC + IW according to the specialties of the word order in Chinese language. For example, in the question of “My grandpa is more than sixty years old and have hypertension, what way can make him better? (爷爷六十几岁了，高血压有什么办法好转?)” the IW is *what* (什么), the NHC is *way* (办法), and the NRC is *hypertension* (高血压). While the question “My grandpa is 73 years old, his blood pressure is 90-160, is it hypertension?(爷爷73岁血压为90-160是高血压吗?)” 高 has the IW (吗) in the last location of the sentence, so it has no NHC, but has a NRC (高血压). The results showed that NRC/VRC + IW worked better than the IW + NHC/VHC which was commonly used for question classification in English.

Table 6 - Studies Related to Automatic Question Classification in Health and Medical Area

Author	Classification task	Machine-Learning Features																
		BOW	n-grams	POS	IW	IW+N HC/V HC	NRC/V RC+I W	concept s	ST	WS	IDF	TF	WL	QL	VS	VO	NE	KTF KIDF KTF-IDF
Cao Y [5]	General topics of clinical Q	✓	✓	✓				✓	✓		✓							
Patrick J [6]	ICU questions	✓	✓		✓	✓			✓	✓					✓	✓		
Liu F [7]	Consumer Q VS professional Q	✓			✓	✓					✓		✓	✓				
Roberts K [8]	Question types for consumer Q	✓		✓	✓	✓		✓			✓		✓				✓	
Conway M [26]	Disease outbreak reports	✓	✓						✓								✓	
Ours	General topics of consumer Q	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓	✓				✓

Note: (1) concepts: UMLS or CMeSH;(2) WS: Wordnet synonyms; (3) VS: Verb-subject relations; (4) VO: Verb-object relations; (5) NE: Named entity based features; (6) KTF: Keywords (TF); (7) KIDF: Keywords (IDF); (8) KTF-IDF: Keywords (TF-IDF).

Lastly, the feature selection methods in our work were quite different from other related works. Cao Y [5], Patrick J [6], Liu F [7], and Roberts K et al [8] employed a method of combining different types of features without considering the threshold, in which they explored different combinations of different feature types and selected the best combination with maximum F₁-score of the classifier. Thus, a feature type would be either picked up or rejected, which may cause the loss of some effective features in the rejected types. Another

disadvantage was the difficulty in exploring all the possible combination of different feature types. Taking a task with 10 types of features for example, it would need to explore 1023 (equal to $2^n - 1$) combinations [27], which was time and resource-consuming. The limited combinations explored in their methods could severely affect the improvement of the classifiers' performance. On the contrary, this research adopted a much more efficient method to combine all the discriminative features from each feature type with $\phi \geq \bar{\phi}$,

which was also proved to be very effective (as described in the *Results* section).

Since the methods proposed in this paper and those in the related works were experimented on different datasets in different languages, it might be inexact to compare the effectiveness of these methods, although the performances of the classifiers trained by our study were quite satisfying.

Limitations

One of the limitations of this study was that the corpus we used to train the classifiers was come from only one Chinese health website, which was focused on hypertension related questions. Therefore, the validity of these classifiers for the vast majority of questions from other websites and other diseases were remained to be tested. Another limitations of this work is that we only reached moderate performances on the classifiers for the topics of *Treatment* and *Epidemiology*. Whereas the error analysis was missed due to the space limitations of this paper.

Conclusion

This study presented a method to automatically identify the topics of consumer health questions in Chinese, so as to facilitate users' information needs extraction and answer retrieval. We explored an abundant feature types and adopted a novel method to select all the effective features with $\Phi \geq \bar{\Phi}$. The performance of the classifiers trained was quite satisfying.

Acknowledgements

This study was supported by Chinese Academy of Medical Sciences (Grant No. 2016ZX330011) and the National Science Foundation of China (Grant No. 14BTQ032). The authors would like to thank Dr. Chao Xu for his help on data processing.

References

- [1] Mao X.L., LI X.M. A survey on question and answering systems. *Journal of Frontiers of Computer Science and Technology* **6** (2012), 193-207.
- [2] Zhen L.H., Wang X.L., Yang S.C. Overview on question classification in question-answering system. *Journal of Anhui University of Technology (Natural Science)* **32** (2015), 48-54.
- [3] Zhang N., Zhu L.J. A survey of Chinese QA system's question analysis. *Technology Intelligence Engineering*, **01** (2016), 32-42.
- [4] Gao Y.Y. Question classification in Chinese question answering system. Hefei: Hefei University of Technology, 2011, 5-14.
- [5] Cao Y., Cimino, J.J., Ely J., et al. Automatically extracting information needs from complex clinical questions. *Journal of Biomedical Informatics* **43** (2010), 962-971.
- [6] Patrick J, Li M. An ontology for clinical questions about the contents of patient notes. *Journal of Biomedical Informatics* **45** (2012), 292-306.
- [7] Liu F., Antieau, L.D., Yu H. Toward automated consumer question answering: automatically separating consumer questions from professional questions in the healthcare domain. *Journal of Biomedical Informatics* **44** (2011), 1032-1038.
- [8] Roberts, K., Kilicoglu, H., Fiszman, M., et al. Automatically classifying question types for consumer health questions. *AMIA Annual Symposium Proceedings. American Medical Informatics Association*, 2014: 1018.
- [9] Liu L., Yu Z.T., Guo J.Y., et al. Chinese question classification based on question property kernel. *International Journal of Machine Learning and Cybernetics* **5** (2014), 713-720.
- [10] Lin X.D., Liu H.Z., Lin P.Y., et al. Chinese question classification using alternating and iterative one-against-one algorithm. *Journal of Convergence Information Technology* **5** (2010), 61-67.
- [11] Zhang D., Li S.S., Wang J.J. Semi-supervised question classification with jointly learning question and answer representations. *Journal of Chinese Information Processing* **31** (2017), 1-7.
- [12] Mollá D., Vicedo, J.L. Question answering in restricted do-mains: an overview. *Computational Linguistics* **33** (2007), 41-61.
- [13] Guo H.H., Li J., Dai T. Consumer health information needs and question classification: analysis of hypertension related questions asked by consumers on a Chinese health website. *Studies in Health Technology and Informatics* **216** (2015), 810-814.
- [14] Li J. Introduction of Rwordseg [EB/OL]. [2016-06-12]. <http://download.csdn.net/detail/c11143015961/8436741>.
- [15] ICTCLAS: Institute of Computing Technology, Chinese Lexical Analysis System [EB/OL]. <http://sewm.pku.edu.cn/QA/reference/ICTCLAS/FreeICTCLAS/>. Accessed June 12, 2016.
- [16] Stanford Log-linear Part-Of-Speech Tagger [EB/OL]. <http://nlp.stanford.edu/downloads/lex-parser.shtml>. Accessed May 16, 2014.
- [17] Interrogative words [EB/OL]. <http://baike.baidu.com/view/720086.htm>. Accessed Oct 21, 2015.
- [18] Niu YQ, Chen J, Duan LG, et al. Study on classification features of Chinese interrogatives. *Computer Application and Software*, 2012, 29(3): 108-111.
- [19] Sun H.X., Li J.L., Li D.Y., et al. The study on semantic mapping from free word to subject headings based on semantic system of CMeSH. *New Technology of Library and Information Service* **29** (2013), 46-51.
- [20] Chen Y.W., Lin C.J. Combining SVMs with various feature selection strategies. *Studies in Fuzziness and Soft Computing* **207** (2006), 315-324.
- [21] Yang L.C., Li J.L., Yao Y.C., et al. A P300 detection algorithm based on F-score feature selection and support vector machines. *Journal of Biomedical Engineering* **25** (2008), 23-26.
- [22] Tsoumakas G., Katakis I. Multi-label classification: an over-view [EB/OL]. <http://ipis.csd.auth.gr/publications/tsoumakas-ijdwm.pdf>. Accessed Nov 15, 2016.
- [23] Gou B, Huang XW. SVM multi-class classification. *Journal of Data Acquisition & Processing* **21** (2006), 334-339.
- [24] The R Project for Statistical Computing [EB/OL]. <https://www.r-project.org/>. Accessed Mar 01, 2016.
- [25] Cross-validation (statistics) [EB/OL]. [https://en.wikipedia.org/wiki/Cross-validation_\(statistics\)](https://en.wikipedia.org/wiki/Cross-validation_(statistics)). Accessed May 16, 2016.
- [26] Conway M., Doan S., Kawazoe A., et al. Classifying disease outbreak reports using n-grams and semantic features. *International Journal of Medical Informatics* **78** (2009), e47-e58.
- [27] Mao S.S., Cheng Y.M., Pu X.L. Probability and mathematical statistics. Beijing: Higher Education Press, 2004: 12-27.

Address for correspondence

Jiao Li, Institute of Medical Information & Library, Chinese Academy of Medical Sciences, 3rd Yabao Road, Chaoyang District, Beijing 100020, China. Email: li.jiao@imicams.ac.cn

Detecting Protected Health Information in Heterogeneous Clinical Notes

Aron Henriksson^a, Maria Kvist^{a,b}, Hercules Dalianis^a

^aDepartment of Computer and Systems Sciences, (DSV), Stockholm University, Sweden

^bDepartment of Laboratory Medicine, Karolinska Institutet, Sweden

Abstract

To enable secondary use of healthcare data in a privacy-preserving manner, there is a need for methods capable of automatically identifying protected health information (PHI) in clinical text. To that end, learning predictive models from labeled examples has emerged as a promising alternative to rule-based systems. However, little is known about differences with respect to PHI prevalence in different types of clinical notes and how potential domain differences may affect the performance of predictive models trained on one particular type of note and applied to another. In this study, we analyze the performance of a predictive model trained on an existing PHI corpus of Swedish clinical notes and applied to a variety of clinical notes: written (i) in different clinical specialties, (ii) under different headings, and (iii) by persons in different professions. The results indicate that domain adaption is needed for effective detection of PHI in heterogeneous clinical notes.

Keywords:

Natural Language Processing, Electronic Health Records, Data Anonymization

Introduction

Healthcare documentation has effectively been digitized as a result of the widespread adoption of electronic health record (EHR) systems. This has, in turn, enabled enormous amounts of valuable data — describing health conditions, treatments and their effects over time — to be analyzed by means of computational methods. To that end, research on developing methods for large-scale analysis of EHR data has surged in recent years, with the promise of improving the effectiveness and efficiency of healthcare, as well as supporting medical research and epidemiological activities [1,2]. However, in order to facilitate the secondary use of EHR data, there is a need for privacy-preserving measures, in particular ones capable of obscuring the sensitive information that is often prevalent in free-text clinical notes. In the absence of informed consent, de-identification is indeed a prerequisite for the secondary use of EHR data: in the U.S., the Health Insurance Portability and Accountability Act (HIPAA) defines 18 types of protected health information (PHI) that should be obscured for EHR data to be considered de-identified [3].

Approaches to automatic de-identification of clinical text can be categorized broadly into rule-based methods and those that employ machine learning [4]. Rule-based methods rely on heuristic rules defined by domain experts, often in conjunction with dictionaries. While rule-based systems sometimes perform well, they are often cumbersome to develop and tend to generalize poorly to new datasets. The machine learning approach is typically supervised, which means that the

learning algorithm relies on manually labeled examples — in this case of PHI — for constructing a predictive model. There are pros and cons of each approach: rule-based systems tend to perform better on rare PHI, while machine learning systems are better able to identify PHI types that are not covered by the employed dictionaries [4]. In the 2006 i2b2 de-identification challenge, machine learning systems performed best [5]. In recent years, the predictive performance of these methods has been improved by, for instance, enriching the feature space describing the data with word embeddings [6,7] or employing alternative learning algorithms, such as deep neural networks [8].

A bottleneck of supervised machine learning is the need for access to labeled examples. In order for predictive models to learn to identify rare classes, in particular, large amounts of data often need to be annotated by domain experts. In the healthcare domain, such resources are scarce and particularly expensive to create. The problem is further exacerbated by the need to cater not only to a multitude of human languages, but also to various sub-domains: performance tends to drop when a predictive model is applied to out-of-domain data, i.e. when the target data is sampled from a different distribution than the training data. In EHRs, there are potentially large domain differences that may need to be accounted for when leveraging predictive modeling to, for instance, automatically identify PHI in clinical text. There are domain differences with respect to various clinical specialties and types of notes, which have been found to be reflected in the vocabulary and linguistic structure of the narrative documentation [9,10]. Different types of clinical notes, moreover, adhere to various levels of formality: discharge notes, for instance, tend to be more formal and function as letters to the general practitioner and other caregivers, while, for example, daily notes are less formal in nature and are primarily intended for the caregiving team, describing health status and progress, and are sometimes also used as memory notes in the ongoing healthcare process. There are also likely to be differences with respect to the author of a clinical note, not only in terms of individual styles and idiosyncrasies but depending on the author's profession, for instance whether it is written by a physician or a nurse. The fact that the content, tone and function of clinical notes often differ substantially may have an impact on the prevalence and expression of PHI, about which very little is known. In a recent study, however, we observed distinct differences in PHI density in clinical notes written by practitioners in different professions, in different specialties, and under different headings [11]. The aim of this study is to investigate whether domain differences in health records may influence the ability of a predictive model, trained on a mixed corpus, to detect PHI in heterogeneous clinical notes.

Methods

In this study, a pre-existing PHI corpus is used for training a predictive model that is subsequently evaluated in nine sub-corpora corresponding to different types of clinical notes. We investigate the prevalence and distribution of PHI in the nine sub-corpora, comprising manual annotations of PHI, and evaluate the ability of the predictive model to detect PHI in the potentially heterogeneous clinical notes. The clinical notes, both in the pre-existing PHI corpus and the newly annotated sub-corpora, are part of the Stockholm EPR Corpus¹, which contains health records in Swedish from Karolinska University Hospital. Names and other sensitive information has been removed from the structured data fields; however, PHI is present in the free-text clinical notes, which may include personal data about the patient, family members, as well as other caregivers in the healthcare process.

The pre-existing PHI corpus comprises 100 health records, in the form of clinical notes, produced in five clinics in 2008: Neurology, Orthopaedics, Infection, Dental Surgery and Nutrition. The clinical notes were manually annotated by three annotators for a set of 28 PHI classes based on HIPAA definitions and enriched with additional, more fine-grained classes — see [12] for more details on the corpus creation process. A consensus-based gold standard was later derived from the original annotations after discussions between the annotators [13]. This process included merging conceptually similar classes, resulting in the following eight classes: *First Name*, *Last Name*, *Age*, *Health Care Unit*, *Location*, *Full Date*, *Date Part* and *Phone Number*. The version of the corpus used in this study contains a total of 198,466 tokens and 4,220 annotated PHI instances. The PHI density — calculated as the number of PHI instances divided by the total number of tokens — is 2.13% and the class distribution is as follows: *Health Care Unit* (23.9%), *First Name* (21.7%), *Last Name* (21.5%), *Date Part* (16.6%), *Full Date* (8.7%), *Location* (3.3%), *Phone Number* (3.2%), *Age* (1.2%).

This latter version of the corpus is then used for training a predictive model. To that end, a linear-chain conditional random fields (CRF) [14] is used that, in addition to being dependent on the input features, is dependent on the previous and subsequent output variables. The following features are used for training and applying the CRF: (a) token, (b) lemma, (c) part of speech, (d) capitalization, (e) digit, (f) compounds, (g) dictionary matching against SNOMED CT, MeSH etc. The same features have been used in previous studies on named entity recognition in Swedish clinical text [15,16] (see [15] for more details). As is common for sequence labeling tasks — in order to account for entity mentions that span multiple tokens — IOB-encoding of class labels is used, which indicates whether a token is at the beginning (B), inside (I) or outside (O) a given named entity mention. 10-fold cross-validation is carried out when tuning the following CRF hyperparameters in a grid-search fashion: two forms of regularization (L1/L2), the c-value governing the balancing between underfitting and overfitting, and the window size, which determines to what extent dependencies should be modeled between input features and output variables. Considered c-values are 2^x , where $x \in \{-2, -1, 0, 1, 2, 3, 4, 5\}$, while the following symmetric window sizes are explored: 1+1, 2+2, 3+3, 4+4. A CRF model is then trained on the entire corpus using the best-observed set of hyperparameters.

The tuned CRF model is then applied on data extracted from 2009 in order to evaluate the predictive performance on

different types of clinical notes. All clinical notes produced in 2009 — i.e., a different year in order to avoid any overlap with the training data and potential bias in the obtained performance estimates — are considered and grouped into three categories according to three dimensions, resulting in a total of nine sub-corpora:

- a) **Speciality** — clinical notes produced in different types of clinics: *Geriatrics*, *Oncology* and *Orthopaedics* (including surgery)
- b) **Heading** — clinical notes written under different headings: *Admission*, *Day*, *Discharge*. Various headings belonging to these general categories are grouped together.
- c) **Profession** — clinical notes written by clinicians in different professional roles: *Physician*, *Nurse*, *Physiotherapist*.

It should be noted that when categorizing notes within one particular dimension, the other dimensions are ignored, which means that a note may exist in more than one sub-corpus, but not across sub-corpora in the same dimension. In order to enable evaluation of predictive performance, subsets of the data are manually annotated. From each sub-corpus, 750 sentences are randomly sampled and annotated by a senior physician using the BRAT annotation tool [17]. The definitions of the eight PHI classes used for the annotation of the corpus that is here used for training the CRF model [12,13] are adhered to; however, elaborate annotation guidelines are not available. For the manual annotation in this study, a number of rules-of-thumb are followed, such as selecting the largest possible scope describing a PHI and not permitting nested or split annotations. *Full Date* is moreover defined to contain year, month and day, while *Date Part* does not comprise all three elements. Age is annotated together with information indicating its status as an age and not just a number, e.g. “A [46-year-old] man”. Names, either *First Name* or *Last Name*, are annotated together as a single entity mention when connected by a hyphen; if not, they are treated as distinct names.

The annotated corpus serves two purposes in the present study: (i) to investigate potential domain differences with respect to the prevalence and distribution of PHI, and (ii) to evaluate and compare the predictive performance of a model trained on a mixed corpus on different types of clinical notes with respect to the three dimensions mentioned above.

Results

When constructing the predictive model, the grid search for the optimal CRF hyperparameters led to the following choices: L2 regularization with a c value of 16 and a narrow (1+1) context window. The predictive performance estimates obtained through 10-fold cross-validation on the training set yielded a precision of 92.65%, a recall of 81.29% and an F1-score of 0.87.

The distribution of manually annotated PHI classes in the samples drawn from the various sub-corpora is depicted in Figure 1, where the total PHI count is shown above each bar. An inter-dimensional comparison shows that PHI are most prevalent in discharge notes and least prevalent in admission notes, as well as notes written by nurses. Overall, the most prevalent types of PHI are *Health Care Unit*, *Last Name* and *Date Part*. The least prevalent types of PHI are *Phone Number*, *Age* and *Location*.

¹ This research has been approved by the Regional Ethical Review Board in Stockholm (2012/834-31/5).

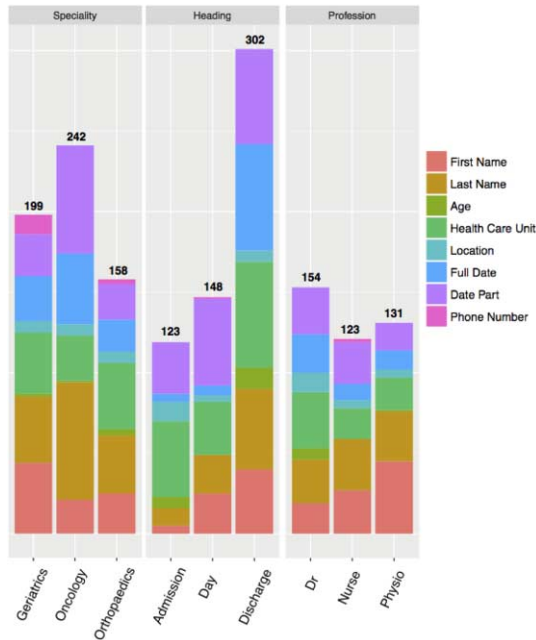


Figure 1 - Prevalence and distribution of PHI in various types of clinical notes

In the speciality dimension, there are more PHI instances in oncology notes (242) compared to geriatrics notes (199) and orthopaedics notes (158). Here, *Last Name* and *Health Care Unit* are generally the most prevalent PHI types, although the proportion of *Health Care Unit* mentions is smaller in oncology notes. Dates, both *Full Date* and *Date Part*, are particularly prevalent in oncology notes, with around 46% of PHI mentions falling into these two categories. *Phone Number* mentions are non-existent in the oncology notes but amount to around 6% of PHI in geriatrics notes.

In the heading dimension, there are considerably more PHI instances — more than double, in fact — in discharge notes (302) compared to daily notes (148) and admission notes (123). Here, the generally most prevalent PHI types are *Health Care Unit* and *Date Part*. In admission notes, the distribution of PHI types is heavily skewed, with *Health Care Unit* (39%) and *Date Part* (27%) representing the largest classes. The PHI distribution is rather less skewed in daily notes and discharge notes. Names, both *First Name* and *Last Name*, are fairly common in these two types of notes, comprising around 30% of PHI mentions, compared to around 13% in admission notes. Dates are also common in all three note types but especially in the latter two. A difference between daily notes and discharge notes is that the latter contains considerably more *Full Date* mentions (22%) than the former (4%), while the former contains more *Date Part* mentions (37% vs. 20%).

In the profession dimension, differences with respect to PHI prevalence is rather less pronounced; however, there are more PHI mentions in physician notes (154) than in physiotherapist notes (131) and nurse notes (123). Here, the generally most prevalent PHI types are *Last Name* and *Health Care Unit*. The distribution of PHI types is most skewed in physiotherapist notes, with names, both *First Name* and *Last Name*, making up around 58% of mentions. Names are also prevalent in nurse notes, a little less so in physician notes but they still make up around 30%. Besides names and mentions of *Health Care Unit*, physicians appear to have a propensity for writing dates — making up around 35% of all PHI mentions — to a larger

extent than do nurses and physiotherapists (less than 30%). Mentions of *Age* are generally rare; however, they do make up 5% of PHI mentions in physician notes.

The predictive performance — in terms of precision, recall and F_1 -score — of the CRF model trained on the pre-existing corpus was then estimated on the nine sub-corpora. Given the rather small sample size, in particular w.r.t. to certain PHI classes, the predictive model is here only evaluated for its ability to detect *any* PHI, which is done by merging the labels post prediction. The performance scores obtained on the three sub-corpora belonging to different specialties are shown in Figure 2. The differences in performance across the three specialties are rather small, with F_1 -scores ranging from around 0.70 to 0.75. The best performance is obtained on oncology notes (F_1 : 0.74), primarily as a result of the relatively high precision (89.93%).

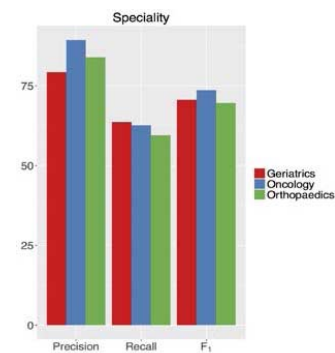


Figure 2 - Predictive performance estimates w.r.t. the ability to detect PHI in clinical notes produced in different types of clinics

The performance scores obtained on the the three sub-corpora comprising clinical notes written under different headings are shown in Figure 3. The predictive performance is generally lower in these sub-corpora, while the differences are somewhat more pronounced: the best performance is obtained on daily notes (F_1 : 0.69) and the worst performance is obtained on admission notes (F_1 : 0.57). The difference between precision and recall is moreover even greater than was observed for clinical notes produced in different clinics: precision is rather high (up to 85.00%, in admission notes), while recall is low (as low as 42.86%, also in admission notes).

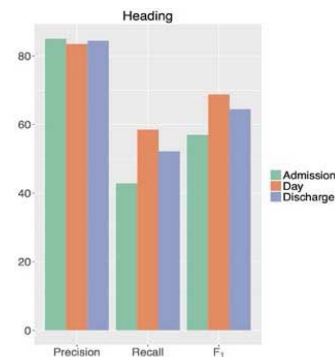


Figure 3 - Predictive performance estimates w.r.t. the ability to detect PHI in clinical notes under different headings

The performance scores obtained on the the three sub-corpora comprising clinical notes written by clinicians in different professional roles are shown in Figure 4. Here, a considerable difference in predictive performance is observed between nurse notes (F_1 : 0.77), on the one hand, and physician notes (F_1 : 0.62) and physiotherapist notes (F_1 : 0.62), on the other.

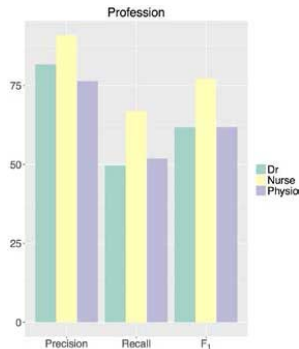


Figure 4 - Predictive performance estimates w.r.t. the ability to detect PHI in clinical notes written by different professions

In general, the predictive performance is markedly lower in comparison to the cross-validated estimates obtained on the training corpus. The F_1 -scores range from 0.57 (admission notes) to 0.77 (nurse notes). Precision is invariably higher than recall: precision ranges from 76.40% (physiotherapist notes) to 91.01% (nurse notes), while recall ranges from 42.86% (admission notes) to 66.94% (nurse notes).

Discussion

The manual annotation of PHI revealed considerable differences with respect to PHI prevalence across the various sub-corpora. Insights into the prevalence and distribution of PHI in different types of clinical notes are useful for directing automatic de-identification efforts and can, for instance, be exploited when creating annotated corpora for learning predictive models. In a previous study [11], we estimated PHI density by applying a predictive model to various types of clinical notes, which comes with the advantage of being able to utilize large amounts of unannotated data; the downside of such a method, however, is that classification errors are not directly taken into account. Manual annotation, as was conducted in the present study, is less prone to errors but requires considerable effort in order to obtain sufficiently large samples. The two methods can, to some extent, be validated by studying to what extent they agree with each other. The results obtained in these two studies show that the methods are generally in agreement. For instance, discharge notes exhibited the highest PHI density² according to both methods, which also revealed high PHI densities in geriatrics and oncology notes. The lowest PHI density was observed in admission notes according to both methods. There are of course also disagreements, e.g. concerning the prevalence of PHI in nurse notes. Many of the observations made in this study with respect to intra-corpus PHI class distributions are also in agreement with the ones made using the method that estimates PHI density based on larger amounts of (unlabeled) data.

² PHI prevalence is used in the current study since an equal number of sentences is sampled from each sub-corpus; in this case only minor differences w.r.t. the number of tokens exist.

Differences with respect to the prevalence and distribution of PHI can, in part, be explained by domain or genre differences. For instance, the informality of daily notes is expressed through the prevalent use of *First Name* and *Date Part*. In contrast, discharge notes are more formal, which is reflected in the prevalent use of *Last Name* and *Full Date*. The high prevalence of PHI in discharge notes is moreover not surprising given that the healthcare process is in some ways summarized therein, which means that, for instance, health care units and physicians that have been involved tend to be named.

The predictive model dropped in performance, in comparison to cross-validated estimates obtained on the training corpus, when applied to the nine sub-corpora. A performance drop can be expected since the cross-validated estimates obtained after an extensive grid search for optimal model hyperparameter values is likely to be somewhat optimistic. It is also to be expected given differences between training and test data, resulting, for instance, in the system having to deal with out-of-vocabulary words. With that in mind, you would expect slightly better predictive performance to have been obtained on the orthopaedics notes since such notes were also included in the training corpus; however, that was not the case. Nevertheless, the more interesting finding based on these experiments is that the observed predictive performance differed substantially, in some cases, between different types of clinical notes. The most notable instance of this phenomenon occurred between nurse notes, on the one hand, and physician notes and physiotherapist notes, on the other. The relatively high performance obtained on nurse notes can, to some extent, be attributed to their low type-token ratio (0.008) and short sentences (8.1 tokens on average). However, other domain differences must be sought for a fuller explanation of the observed differences in predictive performance.

The drop in predictive performance when training a model on one corpus and applying it to other corpora from different domains calls for domain adaptation methods. In the absence of the resources needed for creating a large and representative corpus of clinical notes manually annotated for PHI, such methods would be critical in avoiding the kind of performance drops observed in this study. The relatively low predictive performance obtained on discharge notes is particularly alarming given the high prevalence of PHI therein.

Conclusions

In this study, the prevalence and distribution of protected health information in various types of clinical notes was investigated, revealing considerable differences, including an especially high prevalence of sensitive information in discharge notes. The ability of a predictive model, trained on data drawn from a number of clinics, to generalize to new domains and accurately detect protected health information in various types of clinical notes was moreover assessed. This experiment demonstrated a substantial drop in predictive performance — as well as considerable differences with respect to diverse and heterogeneous clinical notes — when the model was applied to hitherto unseen types of notes, emphasizing the need for domain adaptation. The insights gained into the effects of domain differences, in particular with respect to the prevalence and distribution of protected health information, on predictive performance on heterogeneous clinical notes are valuable for future development of automatic de-identification methods.

Acknowledgements

This study was funded by the Stockholm County Council through the AVID project.

References

- [1] P.B. Jensen, L.J. Jensen, and S. Brunak. Mining electronic health records: towards better research applications and clinical care. *Nature review genetics* 13: 395-405 (2012).
- [2] T.B. Murdoch and A.S. Detsky. The Inevitable Application of Big Data to Health Care. *Journal of the American Medical Association*, 309:13 (2013).
- [3] CDC: Centers for Disease Control and Prevention. HIPAA Privacy Rule and Public Health: Guidance from CDC and the U.S. Department of Health and Human Services (2003), accessed 2016-11-26.
- [4] S.M. Meystre, F.J. Friedlin, B.R. South, S. Shen, and M.H. Samore. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC Medical Research Methodology* 10:1 (2010), 70.
- [5] Ö Uzuner, Y. Luo, and P. Szolovitz. Evaluating the state-of-the-art in automatic de-identification. *JAMIA* 14:5 (2007), 550-563.
- [6] A. Henriksson, S. Kowalski, and H. Dalianis. Generating Features for Named Entity Recognition by Learning Prototypes in Semantic Space: The Case of De-Identifying Health Records. In Proc. of the IEEE International Conference on Bioinformatics and Biomedicine (2014), 450-457.
- [7] A. Henriksson. Learning multiple distributed prototypes of semantic categories for named entity recognition. *Int. J. Data Mining and Bioinformatics* 13: 4 (2015), 395-411.
- [8] F. Demoncourt, J.Y. Lee, P. Szolovits, and Ö. Uzuner. De-identification of Patient Notes with Recurrent Neural Networks. arXiv:1606.03475v1 (2016).
- [9] E. Lövestam, S. Velupillai, and M. Kvist. Abbreviations in Swedish Clinical Text — Use by Three Professions. *Studies in Health Technology and Informatics*, 205: 720-724 (2013).
- [10] K. Smith, B. Megyesi, S. Velupillai, and M. Kvist. Professional language in Swedish clinical text: Linguistic characterization and comparative studies. *Nordic Journal of Linguistics*, 37(02): 297-323 (2014).
- [11] A. Henriksson, M. Kvist, and H. Dalianis. Prevalence Estimation of Protected Health Information in Swedish Clinical Text. *Informatics for Health* (2017).
- [12] H. Dalianis and S. Velupillai. De-identifying Swedish clinical text — refinement of a gold standard and experiments with conditional random fields. *Journal of Biomedical Semantics*, 1:6 (2010).
- [13] S. Velupillai, H. Dalianis, M. Hassel, and G.H. Nilsson. Developing a standard for de-identifying electronic patient records written in Swedish: precision, recall and F-measure in a manual and computerized annotation trial. *International Journal of Medical Informatics*, 78(12): 19-26 (2009).
- [14] J. Lafferty, A. McCallum, and F.C.N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proc. of International Conference on Machine Learning, pp. 282-289 (2001).
- [15] M. Skeppstedt, M. Kvist, G.H. Nilsson, and H. Dalianis. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. *Journal of Biomedical Informatics*, 49: 148-158 (2014).
- [16] A. Henriksson, M. Kvist, H. Dalianis, and M. Duneld. Identifying adverse drug event information in clinical notes with distributional semantic representations of context. *Journal of Biomedical Informatics*, 57: 333-349 (2015).
- [17] P. Stenetorp, S. Pyysalo, G. Topic, T. Ohta, S. Ananiadou, and J. Tsuji. brat: a Web-based Tool for NLP-Assisted Text Annotation. In Proc. of the Demonstrations Session at the European Association of Computational Linguistics (2012).

Address for correspondence

aron.henriksson@dsv.su.se

Acute Coronary Syndrome Risk Prediction Based on GRACE Risk Score

Danqing Hu^a, Zhengxing Huang^a, Tak-Ming Chan^b,
Wei Dong^c, Xudong Lu^a, Huilong Duan^a

^a College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, China,

^b Health Systems, Philips Research China, Shanghai, China,

^c Department of Cardiology, Chinese PLA General Hospital, Beijing, China

Abstract

Clinical risk prediction of acute coronary syndrome (ACS) plays a critical role for clinical decision support, treatment management and quality of care assessment in ACS patients. Admission records contain a wealth of patient information in the early stages of hospitalization, which offers the opportunity to support the ACS risk prediction in a proactive manner. However, ACS patient risks aren't recorded in hospital admission records, thus impeding the construction of supervised risk prediction models. In our study, we propose a novel approach for ACS risk prediction, which employs a well-known ACS risk prediction model (GRACE) as the benchmark methods to stratify patient risks, and then utilizes a state-of-the-art supervised machine learning algorithm to establish our risk prediction models. The experiment was conducted with a collection of 3,643 ACS patient samples from a Chinese hospital. Our best model achieved 0.616 accuracy for risk prediction, which indicates our learned model can achieve a better performance than the benchmark GRACE model and can obtain significant improvement by mixing up patient samples that were manually labeled risks.

Keywords:

Risk Assessment; Acute Coronary Syndrome; Supervised Machine Learning

Introduction

Acute coronary syndrome (ACS) refers to a group of conditions where, due to decreased blood flow in the coronary arteries, part of the heart muscle is unable to function properly or dies [1, 2]. Clinical risk prediction of ACS, as an essential procedure in patient care delivery, plays a critical role for clinical decision support, patient treatment management, as well as the quality of care assessment [1, 3]. For a long time, a lot of studies have focused on this important topic and established a lot of ACS risk prediction tools such as GRACE [4], TIMI [5], Pursuit [6] and Framingham [7], which have illustrated a competitive performance in real clinical scenarios. Although useful, these tools have drawbacks such as limited consideration of risk factors, a long period of data collection, and strict inclusion/exclusion criteria [3].

Nowadays, more and more information systems are implemented in hospitals, which generate a large volume of electronic medical records (EMR) to support the study of medical informatics. In particular, admission records that capture a large amount of patient information in the early stages of hospitalization make it possible as brand-new media for ACS risk prediction. In our previous work [8], we have proposed a Rule-based Medical Language Processing method to extract

features from admission records. However, patient risks are not explicitly recorded in their admission records, which impedes our employment of supervised machine learning algorithms to build our risk prediction models. Therefore, we address this issue in our study by applying a well-known ACS risk prediction model, i.e., GRACE, which was established by the full spectrum of patients, as our benchmark method to stratify each patient sample into 3 risk levels, i.e., low-, medium- and high-risk levels, based on a pre-selected set of risk factors identified in the GRACE model and extracted from the admission records. As such, the corresponding class label, i.e., risk levels of patient samples, can be assigned to that patient. Thereafter, we employ three state-of-the-art supervised machine learning algorithms, i.e., Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression with ℓ_1 regularization on the regression coefficients (ℓ_1 -LR), to build ACS risk prediction models. The experiments have been conducted on a real clinical dataset consisting of 3,463 patient samples collected from the cardiology department of Chinese PLA General Hospital. Experimental results demonstrate the feasibility of our novel approach for ACS risk prediction.

Methods

Data

A total of 3,463 admission records were collected from the Cardiology Department of the China PLA General Hospital, which covers heterogeneous aspects including demographics, chief complaint, medical history, lab test and specific inspection, etc. These data can be used to provide ACS risk prediction service at an early stage of hospitalizations. The average length of stay (LOS) is 8.20 days, while some patients take a very short time, i.e., only 1 day in the hospital, and others take much longer, i.e., more than 3 months in the hospital, which implicitly indicates the diversity of patient conditions in the treatment process. In our previous study, we employed a Rule-Based Medical Language Process (RBMLP) method to extract features from the free text of admission records and a simple regular expression to extract vital signs and lab test results that are in a structured format to support the construction of ACS risk prediction models. A total of 268 features were extracted using the RBMLP which covered 9 different categories. The evaluation of feature extraction results has been implemented depending on manually labeled features of 100 randomly selected patient samples by an experienced clinical engineer. The precision, recall, and F-score are 98.56%, 95.26% and 96.88%, respectively, which indicate that the feature extraction results can effectively reflect the patient conditions. Table 1 shows the summaries of extracted patient features.

The case study was performed in the Cardiology Department at the Chinese PLA General Hospital. Prior approval was obtained from the data protection committee of the hospital to conduct the study. Patient data were anonymized in this study and in this paper.

Table 1 – The summaries of extracted patient features

Vital signs related features	Mean	S.D.	Lab test related features	F.A.	Free-text features	F.A.
BT(°C)	36.2	±0.3	HDL-C	42.5%	CHD	97.7%
Pulse(bpm)	73.4	±9.7	Calcium	34.8%	Angina	94.5%
BR(bpm)	18.1	±0.5	Triglycerides	33.1%	Arteriosclerosis	77.0%
SBP(mmHg)	132.2	±17.9	Glucose	29.9%	CPS	74.1%
DBP(mmHg)	77.4	±10.4	Neutrophils	29.2%	Hypertension	68.1%
Height(cm)	166.9	±8.1	Eosinophils	28.0%	PRP	52.5%
Weight(kg)	71.5	±12.3	Urine color	27.9%	Anti-H therapy	47.9%
			Lymphocytes	26.4%	Anti-A therapy	42.5%
			Monocytes	21.4%	Anti-C therapy	42.1%
			Apo A-1	18.8%	Sweating	38.2%

S.D.: standard deviation, F.A.: frequency of abnormal values, BT: body temperature, BR: breathe rate, SBP: systolic blood pressure, DBP: diastolic blood pressure, HDL-C: high density lipoprotein-cholesterol, Apo A-1: apolipoprotein A-1, CHD: Coronary Heart Disease, CPS: chest pain and stuffiness, PRP: precordial region pain, Anti-H therapy: anti-hypertension therapy, Anti-A therapy: anti-angina therapy, Anti-C therapy: anticoagulant therapy

Benchmark dataset

While an admission record contains valuable patient information like demographics, medical history, lab test and specific inspection, and diagnosis codes, etc., the patient's risk score, as an important clinical outcome indicator in ACS treatment processes and necessary information for constructing training samples of ACS risk prediction, is not directly available in the records. To address this problem, we applied a well-known ACS risk scoring model, i.e., GRACE [4], as the benchmark method to determine the risk score of an ACS patient by referring to his or her admission record. GRACE is an international observational database designed to reflect data from an unbiased population of ACS patients. It is established by the full spectrum of ACS patients (ST-segment elevation myocardial infarction, non-ST-elevation myocardial infarction, and unstable angina). It has been proved that the model can provide excellent ability to assess the patient's risk [9]. In particular, we stratified each ACS patient sample into one of three risk levels (i.e., low-, medium-, and high-risk levels) based on the calculated GRACE score [10].

Note that GRACE scores are not presented in our original dataset, and were calculated on a pre-selected set of risk factors extracted from admission records by the proposed methods presented in our previous work. However, missing data is a problem intrinsic in free text information extraction and our work is no exception [11]. While the most common way to handle the missing data is to exclude or discard the data, it is not suitable for our work because of the amount of missing value of our data, and it may also create biased estimates [12]. To overcome the limitations, we employed three strategies based on the different risk factors in the GRACE model. Table 2 summarizes the strategies we employed for each risk factor. For Heart rate and Systolic blood pressure, we replaced the missing values with mean values (Strategy 1), and we employed multiple imputations [13] (Strategy 2) to impute the missing values in creatinine and elevated cardiac markers, and for the ST-segment deviation, Killip class and cardiac arrest at admission, we considered the patients were normal if we didn't extract the information indicating their abnormalities (Strategy 3). While the missing value percentages for Strategy 3 risk factors seem to be huge, it is effective as physicians recorded only the activities deviating from normal (charting by exception). Based on the strategies to handle missing values, we obtained the completed imputed dataset. Accordingly, the GRACE model can be employed to label the corresponding risk classes to patients.

Table 2 – Imputation strategies for GRACE risk Factors

GRACE Risk Factor	Missing value (%)	Strategies
Age	0%	N/A
Heart rate	0.06%	Strategy 1
Systolic BP	0.23%	Strategy 1
Creatinine	3.93%	Strategy 2
Elevated cardiac markers	2.86%	Strategy 2
Killip class	91.22%	Strategy 3
ST-segment deviation	81.52%	Strategy 3
Cardiac arrest at admission	99.97%	Strategy 3

Strategy 1: Replace the missing values with mean values Strategy 2: Multiple Imputations Strategy 3: Consider the patient is normal if there is no extracted information indicates his/her abnormalities.

Risk prediction model

Based on steps above, we can obtain a set of patient features from admission records and their corresponding risk levels based on GRACE. We can then build our risk prediction models using well-established machine learning algorithms with their predictive accuracy on the hold-out samples. In this study, three representative supervised machine learning classifiers, including Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression with ℓ_1 regularization on the regression coefficients (ℓ_1 -LR), were selected and compared to each other due to their demonstrated predictive performance and their wide usage in the recently published research literature. All classifiers were multinomial models (eligible for three classes: low-risk, medium-risk and high-risk). A 5-fold cross validation was performed on each classifier. We repeated the learning process 10 times to measure the average performance and the 95% confidence interval of each classifier. What follows is a brief introduction to the employed algorithms as clinical risk prediction tools:

- SVM is one of the most popular classification models based on constructing a hyper-plane or set of hyperplanes in a high- or infinite-dimensional space to classify [14]. In this study, RBF kernel was employed to build the non-linear SVM classifier with parameters, i.e., Cost and Gamma, determined by cross validation [15].
- RF is an ensemble learning method that operates by constructing a multitude of decision trees and outputting the class that is the mode of the classes of the individual trees [16]. In our study, we set 1000 as the number of trees in the forest and the square root of the number of features as the number tried at each split [16].
- ℓ_1 -LR is the LR with a lasso penalty factors on the regression coefficients [17]. ℓ_1 -LR has been widely used for many classification problems, particularly ones with many features. It is well-known that regularization is required to avoid over-fitting, especially when there are a large number of parameters to be learned and can be used as feature selection methods [17]. In our study, we determined the parameter, i.e., Lambda, depending on cross validation.

To evaluate the performance of the learned models, we employed the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) to evaluate the learning algorithms. Note that the ROC curve can only evaluate the performance of binary classifiers, we plot three ROC curves, i.e., low-risk or not, medium-risk or not and high-risk or not, for each model using the one-vs-rest approach. We also calculated the micro- and macro-averaged ROC curve for each model to show the overall performance. Moreover, we calculated the prediction accuracies for each model and employed Wilcoxon signed-rank test to compare the differences between GRACE and the learned models. All

model constructions and statistical analyses were completed using R version 3.2.1 (R Foundation for Statistical Computing) and Scikit-learn library [18] under Python version 3.5.1.

Experiments and Results

Benchmark dataset construction

According to our previous work, we have extracted patient features from their admission records using our RBMLP method. The summaries of extracted features are illustrated in Table 1. Based on the feature extraction results, we employed the well-known ACS risk prediction model, GRACE, as the benchmark method to label the risk levels to the patients in the data pool. In particular, we designed three different imputation strategies to overcome the missing value limitation. The summaries of labeled patient samples were illustrated in Table 3. After the benchmark dataset construction, 1,968, 1,054 and 441 patients were stratified into low-, medium- and high-risk levels, respectively.

Table 3 – The summaries of labeled patient samples

Low risk samples	Medium risk samples	High risk samples
1,968	1,054	441
GRACE Items	Risk Scores	Frequency (%)
Age (year)		
<40	0	68 (1.96%)
40-49	18	443 (12.79%)
50-59	36	908 (26.22%)
60-69	55	866 (25.01%)
70-79	73	842 (24.31%)
≥80	91	336 (9.70%)
<40	0	68 (1.96%)
Heart rate (bpm)		
<70	0	1,122 (32.40%)
70-89	7	2,164 (62.49%)
90-109	13	164 (4.74%)
110-149	23	13 (0.38%)
150-199	36	0 (0%)
≥200	46	0 (0%)
Systolic BP (mmHg)		
<80	63	1 (0.03%)
80-99	58	35 (1.01%)
100-119	47	439 (12.68%)
120-139	37	1,897 (54.78%)
140-159	26	786 (22.70%)
160-199	11	299 (8.63%)
≥200	0	6 (0.17%)
Creatinine (μmol/L)		
0-35	2	11 (0.32%)
36-70	5	1,400 (40.43%)
71-105	8	1,706 (49.26%)
106-140	11	223 (6.44%)
141-176	14	60 (1.73%)
177-353	23	45 (1.30%)
≥354	31	18 (0.52%)
Elevated cardiac markers	15	648 (18.71%)
Killip class		
Class I	0	3,230 (93.27%)
Class II	21	134 (3.87%)
Class III	43	88 (2.54%)
Class IV	64	11 (0.32%)
Cardiac arrest at admission	43	1 (0.03%)
ST-segment deviation	30	569 (16.43%)

Risk prediction model construction

We employed three representative supervised classification methods (i.e., RF, SVM and ℓ_1 -LR) described above under ‘Risk prediction model’, with datasets constructed by the RBMLP feature extraction methods in our previous work. And then, we employed metrics of ROC and AUC to evaluate the learned models. Considering that the ROC curve can only evaluate the performance of the binary classifier, we plot three ROC curves (low-risk or not, medium-risk or not and high-risk or not) for each model respectively. Then we integrated the model’s three curves into Micro- and Macro-Averaged ROC

curve [18] to show the overall performance of the model. Furthermore, to analyze the influences of changes in sample sizes on the model’s AUC, we randomly selected the same proportion of samples from three risk levels. By the 10 times repetition of learning process under 10 different sample sizes, we averaged the 10 generated curves’ AUC under each sample size to form a curve with 95% confidence interval (CI). The left side of Figure 1 shows the ROC curves of each model and the right illustrates the AUC changes under different sample sizes. All three learned models achieved acceptable performance using the GRACE-labeled dataset, which indicates that our models obtained a good fitting degree with the GRACE model. ℓ_1 -LR achieved best performances (0.98 in Micro-averaged ROC and 0.97 in Macro-averaged ROC). And moreover, the AUC of all learned models gradually ascended and became more stable along with the increase of sample size, which indicates that our learned models can achieve more competitive performance under a larger sample size.

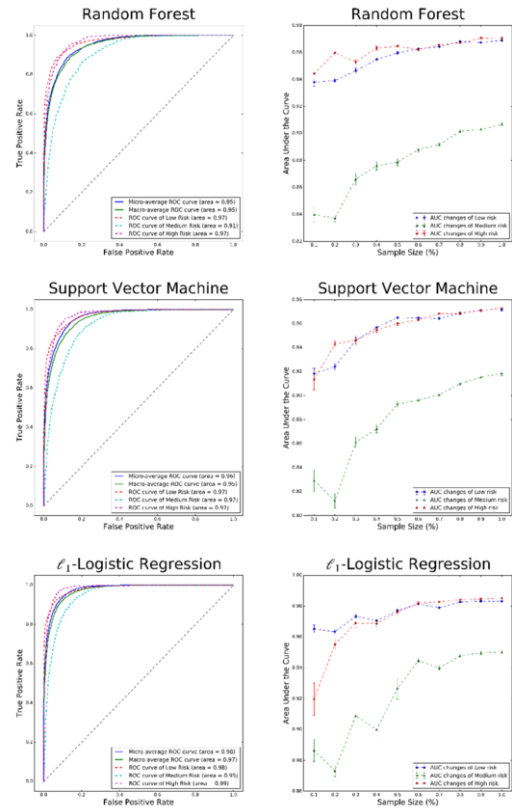


Figure 1 – ROC curves and AUC changes of learned models

Risk prediction model evaluation

As presented above under ‘Risk prediction model construction’, we have shown that our models can achieve a very high fitting degree with the GRACE model. In this section, we further evaluate the effectiveness of our models on ACS risk prediction, in comparison with GRACE via a small set of manually labeled ACS patient samples. As mentioned above, the risk labels of the training dataset are provided by the GRACE model. Although useful, it might introduce errors and noises during the risk labeling process via GRACE since the actual outcome of an ACS patient is unknown. A more reasonable method is to review an ACS patient’s progress notes in the hospital to check the outcomes of that patient. As such,

the actual adverse event label can be provided for that patient. Thus, it raises the question if we can learn a more effective and robust risk prediction model from an accurately labeled dataset. To this end, we randomly selected 500 patient samples from all 3,463 original patient records and then asked an experienced clinical engineer to manually provide risk labels of these patients by reviewing their progress notes in their hospitalizations. Then, we divided the manually labeled samples into 5 parts and each part was predicted by the model retrained based on the rest 4 parts and the other 2,963 (3,463-500) GRACE-labeled patient samples. To evaluate the performance of the benchmark GRACE model, our risk prediction models learned from both the GRACE-Labeled dataset (namely Models-GL), and the dataset mixed with Partly Manually Labeled patient samples (namely Models-PML). The accuracy of each model is computed and shown in Figure 2. In addition, we employed Wilcoxon signed-rank test to check the significant differences between the GRACE and the learned models. The experimental results are shown in Table 4.

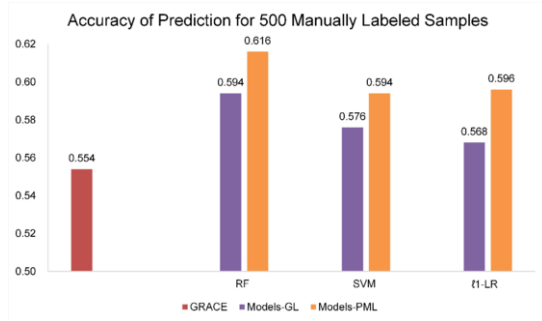


Figure 2 – Accuracy of each model for predicting the 500 manually labeled patient samples

Table 4 – Wilcoxon signed ranks test between GRACE, Models-GL and Models-PML

Models	RF	SVM	t1-LR
GRACE VS. Models-GL	1.51×10^{-7} ***	2.33×10^{-1} •	4.17×10^{-1} •
GRACE VS. Models-PML	1.06×10^{-11} ***	4.14×10^{-4} ***	2.07×10^{-7} ***
Models-GL VS. Models-PML	1.05×10^{-5} ***	2.74×10^{-5} ***	4.62×10^{-10} ***

***: p-value < 0.005; **: p-value < 0.01; *: p-value < 0.05; •: p-value > 0.05

As shown in Figure 2, the GRACE model achieves 55.4% accuracy for the 500 manually labeled patient samples. It indicates that the GRACE model has a good discriminatory ability in predicting ACS patients' risks. More surprisingly, all our models, i.e., RF, SVM and t1-LR models, learned from the GL dataset and the PLM dataset to achieve competitive performance. In Figure 2, the prediction accuracy of the best model (i.e., RF-PLM) is 61.6%, and is higher than the accuracy obtained by the GRACE model. The Wilcoxon signed-rank test results proved the significant improvements of our models in comparison with GRACE. In addition, we compared the performance obtained from both Models-GL and Models-PML in terms of accuracy, to explore whether the prediction performance can be significantly improved from the manually labeled samples. As shown in Table 4, there are significant differences of Models-GL and Models-PML, in which we argue that the proposed approach can obtain better prediction performance from manually labeled patient samples.

Discussion

The ability to analyze a large volume of EMR data in modern healthcare systems plays a vital role in the improvement of the quality of care delivery [19]. In this paper, we employed a well-known ACS risk prediction model, i.e., GRACE, as our benchmark method to automatically stratify 3,463 patient samples into low-, medium- and high-risk levels, respectively. In particular, three imputation strategies were presented to solve the missing value limitation in our dataset. Then, three competitive supervised machine learning algorithms were selected to construct our risk prediction models.

Experiments have been conducted in a total of 3,463 ACS patient samples collected from a Chinese hospital to evaluate the performance of our models. By the evaluation of 500 manually labeled samples, our Models-GL achieved better performances than the GRACE model. The Models-PML obtained a significant improvement in comparison with GRACE and Models-GL, which indicates that the risk model's performance can be elevated by the mixing of manually labeled patient samples.

However, there are still certain limitations in our current work. In this study, the labels in the training dataset are provided by the benchmark GRACE model. Note that, GRACE, as a prospectively studied scoring system, is not always effective in risk prediction in various clinical scenarios [20]. The intrinsic problem of free-text-extracted data, i.e., missing values, also made the labels that were provided by benchmark GRACE model less accurate, though we have employed imputation strategies to handle this problem. As indicated in our experimental results, incorporating manually labeled patient samples provided by experimented physicians into the training dataset can improve the performance of our models in ACS risk prediction. Thus, the better way to build a benchmark dataset is to retrospectively study medical cases to stratify ACS patient's risks by experienced physicians. However, it is a tedious and time-consuming process for human evaluation on a large volume of patient samples. To this end, semi-supervised learning could provide a feasible solution since it can use ready-to-use unlabeled data to improve supervised learning tasks when the labeled data is scarce or expensive [21].

Conclusion

In this paper, we have presented a novel approach to utilize the well-known ACS risk prediction model, GRACE, as the benchmark for the establishment of improved models for ACS risk prediction. The experimental results show that our learned models can achieve a better performance than the GRACE model and obtain a significant improvement by mixing manually labeled patient samples into the training dataset, which illustrates a brand-new approach to risk prediction model construction using existing methods.

Acknowledgements

This work was supported by the National Nature Science Foundation of China under Grant No. 61672450, and the Brain-Bridge Project: Risk assessment to improve patient outcomes for cardiovascular diseases, from Philips (China) Investment Co., Ltd. (Philips), Zhejiang University and Eindhoven University of Technology. The authors are especially thankful for the positive support received from the cooperative hospitals as well as to all medical staff involved.

References

- [1] M.T. Cooney, A.L. Dudina, and I.M. Graham, Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians, *J Am Coll Cardiol* **54** (2009), 1209-1227.
- [2] Acute coronary syndrome. https://en.wikipedia.org/wiki/Acute_coronary_syndrome, Last access on 2015-10-24.
- [3] S. Paredes, T. Rocha, P. de Carvalho, J. Henriques, M. Harris, and J. Morais, Long term cardiovascular risk models' combination, *Comput Methods Programs Biomed* **101** (2011), 231-242.
- [4] C.B. Granger, R.J. Goldberg, O. Dabbous, K.S. Pieper, K.A. Eagle, C.P. Cannon, F. Van De Werf, A. Avezum, S.G. Goodman, M.D. Flather, K.A. Fox, and G.R.A.C.E. Investigators, Predictors of hospital mortality in the global registry of acute coronary events, *Arch Intern Med* **163** (2003), 2345-2353.
- [5] E.M. Antman, M. Cohen, P.J. Bernink, C.H. McCabe, T. Horacek, G. Papuchis, B. Mautner, R. Corbalan, D. Radley, and E. Braunwald, The TIMI risk score for unstable angina/non-ST elevation MI: A method for prognostication and therapeutic decision making, *JAMA* **284** (2000), 835-842.
- [6] E. Boersma, K.S. Pieper, E.W. Steyerberg, R.G. Wilcox, W.C. Chang, K.L. Lee, K.M. Akkerhuis, R.A. Harrington, J.W. Deckers, P.W. Armstrong, A.M. Lincoff, R.M. Califf, E.J. Topol, and M.L. Simoons, Predictors of outcome in patients with acute coronary syndromes without persistent ST-segment elevation. Results from an international trial of 9461 patients. The PURSUIT Investigators, *Circulation* **101** (2000), 2557-2567.
- [7] S. Quaglini, M. Stefanelli, L. Boiocchi, F. Campari, A. Cavallini, and G. Miceli, Cardiovascular risk calculators: understanding differences and realising economic implications, *Int J Med Inform* **74** (2005), 191-199.
- [8] D. Hu, Z. Huang, T.-M. Chan, W. Dong, X. Lu, and H. Duan, Utilizing Chinese Admission Records for MACE Prediction of Acute Coronary Syndrome, *International Journal of Environmental Research and Public Health* **13** (2016), 912.
- [9] P. de Araújo Gonçalves, J. Ferreira, C. Aguiar, and R. Seabra-Gomes, TIMI, PURSUIT, and GRACE risk scores: sustained prognostic value and interaction with revascularization in NSTEMI-ACS, *Eur Heart J* **26** (2005), 865-872.
- [10] C.W. Hamm, J.P. Bassand, S. Agewall, J. Bax, E. Boersma, H. Bueno, P. Caso, D. Dudek, S. Gielen, K. Huber, M. Ohman, M.C. Petrie, F. Sonntag, M.S. Uva, R.F. Storey, W. Wijns, D. Zahger, and E.C.f.P. Guidelines, ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC), *Eur Heart J* **32** (2011), 2999-3054.
- [11] S. Bandyopadhyay, J. Wolfson, D. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. Johnson, and P. O'Connor, Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data, *Data Mining and Knowledge Discovery* **29** (2015), 1033-1069.
- [12] J. Jonnagaddala, S.T. Liaw, P. Ray, M. Kumar, N.W. Chang, and H.J. Dai, Coronary artery disease risk assessment from unstructured electronic health records using text mining, *J Biomed Inform* **58 Suppl** (2015), S203-210.
- [13] S. van Buuren and K. Groothuis-Oudshoorn, mice: Multivariate Imputation by Chained Equations in R, *2011* **45** (2011), 67.
- [14] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning* **20** (1995), 273-297.
- [15] W.H. Press, *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge university press, 2007.
- [16] L. Breiman, Random Forests, *Machine Learning* **45**, 5-32.
- [17] S.-I. Lee, H. Lee, P. Abbeel, and A.Y. Ng, Efficient L1 regularized logistic regression, in: *Proceedings of the 21st national conference on Artificial intelligence - Volume 1*, AAAI Press, Boston, Massachusetts, 2006, pp. 401-408.
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.* **12** (2011), 2825-2830.
- [19] Z. Huang, W. Dong, P. Bath, L. Ji, and H. Duan, On mining latent treatment patterns from electronic medical records, in: *Data Mining and Knowledge Discovery*, 2015, pp. 914-949.
- [20] K.A. Fox, K.A. Eagle, J.M. Gore, P.G. Steg, F.A. Anderson, and G.a.G. Investigators, The Global Registry of Acute Coronary Events, 1999 to 2009--GRACE, *Heart* **96** (2010), 1095-1101.
- [21] X. Zhu and A.B. Goldberg, Introduction to Semi-Supervised Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning* **3** (2009), 1-130.

Address for correspondence

Corresponding author: Zhengxing Huang

Mailing Address: Zhouyiqing Technology Building 512, Zheda Road 38, Zhejiang University, Hangzhou, Zhejiang, China.

Email: zhengxinghuang@zju.edu.cn

Phone: +86 13819488552

Prediction and Factor Extraction of Drug Function by Analyzing Medical Records in Developing Countries

Min Hu^a, Yasunobu Nohara^b, Masafumi Nakamura^a, Naoki Nakashima^b

^a Graduate School of Medical Sciences, Kyushu University, Fukuoka City, Fukuoka, Japan,

^b Medical Information Center, Kyushu University Hospital, Fukuoka City, Fukuoka, Japan,

Abstract

The World Health Organization has declared Bangladesh one of 58 countries facing acute Human Resources for Health (HRH) crisis. Artificial intelligence in healthcare has been shown to be successful for diagnostics. Using machine learning to predict pharmaceutical prescriptions may solve HRH crises. In this study, we investigate a predictive model by analyzing prescription data of 4,543 subjects in Bangladesh. We predict the function of prescribed drugs, comparing three machine-learning approaches. The approaches compare whether a subject shall be prescribed medicine from the 21 most frequently prescribed drug functions. Receiver Operating Characteristics (ROC) were selected as a way to evaluate and assess prediction models. The results show the drug function with the best prediction performance was oral hypoglycemic drugs, which has an average AUC of 0.962. To understand how the variables affect prediction, we conducted factor analysis based on tree-based algorithms and natural language processing techniques.

Keywords:

Data Mining; Health Services Accessibility; Medical Informatics

Introduction

The World Health Organization has declared Bangladesh one of 58 countries facing acute Human Resources for Health (HRH) crisis [6]. Health care provider densities in Bangladesh are only 0.3 physicians and 0.28 nurses per 1,000 people. In comparison, the health care provider densities in Japan are 2.1 physicians and 7.9 nurses per 1,000 people.

Recently, artificial intelligence-based healthcare systems are highly regarded due to their success in diagnosis, prediction, and choice of treatment [3-5]. Karimi et al., [1] used a decision tree and support vector machine (SVM) to predict metabolic syndrome from 2,107 participants. Colubri et al., [2] developed a machine-learning pipeline for Ebola prognosis prediction that packages the best models into a mobile app that is available in clinical care settings.

Kyushu University and Grameen Communications have conducted a study [7] into health management using information communication technology. The Portable Health Clinic (PHC) developed a package with medical sensors (e.g., blood pressure monitors, blood glucose meters, and urine tape) to provide immediate consultation with a remote doctor over Skype. Operators can use the system with minimal information technology literacy to provide health checkup services, even in

rural areas. Following the tele-consultation, the remote doctor can give the patient an electronic prescription.

Data collection from 16,741 subjects was conducted in 10 locations in Bangladesh, including urban and rural regions, between July 2012 and February 2014 [7].

In this study, we investigate a predictive model by analyzing the prescription data from the Health Checkup and Telemedical Intervention program in Bangladesh. We consider the function of prescribed drugs as a diagnosis for subjects. Three-machine learning algorithms, including a Gradient Boosting Decision Tree (GBDT), an SVM, and Logistic Regression were employed to predict whether a subject is prescribed one of several most frequently prescribed drug types. We extract important factors from each medication prediction. To understand better, how the variables affect prediction, we conducted factor analysis using tree-based algorithms and natural language processing techniques. We aim to use the results of this study in the future to implement a web-based application to support clinical decision making relative to medicine prescriptions in developing countries.

Methods

Data collection procedure

PHC comprises a set of sensor devices in an attaché case, a data transmission system linked to a mobile network, and a data management application.

We provided a healthcare service for the study, including health checkups that used sensor devices in the PHC, data storage in a call center, a health report, and healthcare guidance according to the given individual, and a tele-consultation with a doctor in a medical call center. We conducted the study in five rural villages and five factories/offices in Bangladesh.

During the first visit, after registration, the subjects received an ID card with a barcode. After completing a questionnaire, the subject underwent a health checkup using the sensor devices in the PHC. Qualified healthcare professionals performed both blood glucose and urine tests, and trained staff performed other tests.

We crosschecked the urine test results of the workers every two or three months because this test requires a visual assessment. Other devices, including blood glucose devices, display numerical results that do not require calibration among workers. The data were stored on an Android tablet and the main server in Dhaka. We implemented categorized results for four risk groups for PHC, which are green, yellow, orange and red, according to a rule-based triage algorithm, in which, local staff printed and explained to the subject. A booklet was provided to all subjects who are marked for health risk according

to test results. We provided telemedical intervention with a doctor in Dhaka for orange- and red-grade subjects. Because we selected sites from around Dhaka in the first year, subjects in the village and factories/offices around Dhaka were asked to undergo a health checkup one year later to assess the effects of the program.

After the health checkup, we provided telemedical intervention for orange- and red-grade subjects, via mobile network contact (Skype), with the medical call center in Dhaka. Because most areas in Bangladesh have Internet access (2G/3G), we brought laptop PCs or tablet PCs (iPad) with mobile routers to the checkup sites. The staff set up special rooms for teleconsultations at checkup sites and assisted subjects in their language with remote doctors in Dhaka. Doctors had access to the health checkup results via the Internet and could provide advice about disease management and encourage subjects to visit a clinic. Where required, the doctors could send a teleprescription for anti-hypertensive medication via the network. In our program, subjects who received a teleprescription could visit their local pharmacy to purchase medication.

PHC field study was conducted between July 2012 and March 2014 (first year: July 2012–February 2013; second year: June 2013–March 2014). In total, 16,741 subjects were assessed in the first health checkup and 4,543 were provided remote consultations. The remote consultation records, which contain information about chief complaints (CC), doctor advice, and e-prescriptions, were input in Bengali during the field study and then translated to English. Ethics Review Committee of Kyushu University Hospital has approved this study (Ethics Review Number: 24048).

Table 1 shows the classified function of medicine prescribed to more than 100 subjects. We identified 21 pharmaceuticals, and we predicted whether each drug has been prescribed.

Table 1- Types of pharmaceuticals prescribed for more than 100 subjects

Type of prescription (drugs for)	Number of people
Peptic ulcer	1,760
Mineral preparations	709
Analgesics & antipyretics	669
Hypertension	669
Sedation & hypnosis	518
Multivitamin & multi-mineral combined preparations	498
Inflammation and rheumatic diseases	496
Iron deficiency anemias	477
Bacterial infections	413
Specific vitamin preparations	390
Angina & ischemic heart diseases	316
Oral hypoglycemic drugs	276
Allergic disorders	269
Emetics	226
Epilepsy	222
Psychosis	200
Protozoal infections	194
Depression	186
Neuromuscular disorders	141
Asthma & prophylaxis	102
Constipation	102

Drug hierarchical structure

The chemical type of active ingredient may classify a drug or by the way, it is used to treat a particular condition. We use a four-layer hierarchical drug classification system based on a Bangladesh medicine database, which has information in Eng-

lish on Bangladesh medicines [8]. Figure 1 shows the hierarchical drug classification system.

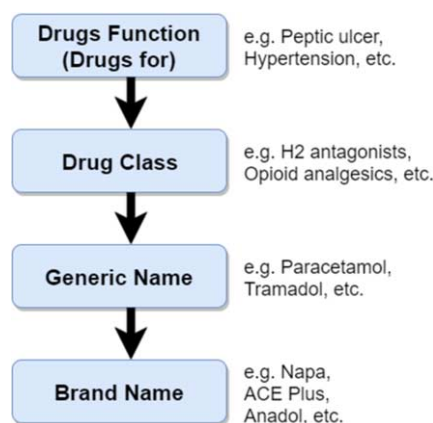


Figure 1- Drug hierarchical classification system of bangladesh medicines

“Drug Function” (Drugs for) and “Drug Class” are the first and second levels of the hierarchical drug classification system. Drugs used for the same purpose belong to the same “Drug Function” block and drugs that work in the same way belong to the same “Drug Class” block. “Generic Name” is the third level and is classified by the chemical type of the active ingredient. Many pharmaceutical companies market generic drugs, and these companies select a “Brand Name” (fourth hierarchy level) for their products.

In the PHC field study, health checkup prescription data is input and recorded as “Brand Name.” We added “Drug Function” information to the data by querying a Bangladesh medicine database using the brand name of each pharmaceutical [8]. Since a doctor’s diagnosis is not contained in the remote consultation record, “Drug Function” information can be considered an alternative way to understand a subject’s health problem.

Explanatory variables

We used 1,550 variables as explanatory variables to predict pharmaceuticals to be prescribed.

Basic information and measurement results (49 variables)

The subject profile contains gender, age, site ID (15 sites), site type (rural, suburban, or urban), and checkup date. Measurement items are waist, hip, waist-hip ratio, height, weight, BMI, systolic blood pressure, diastolic blood pressure, blood sugar (BS), BS type (postprandial or fasting), urine protein, urine sugar, urine urobilinogen, pulse rate, arrhythmia, body temperature and SpO2. Each measurement result is classified (four colors) using specific diagnostic rules we implemented for PHC [7]. These colors are also used as explanatory variables.

Interview sheet (50 variables)

Before receiving remote diagnosis by a doctor, the subjects were interviewed by local staff. The interview was conducted using an interview sheet. The interview sheet contains 31 questions about occupation, present symptoms, past diseases, medication, smoking, weight change, exercise, walking speed, eating behavior, sleeping habits, the desire to have a healthy lifestyle, drug allergies, and surgical history. Since some questions are given in a multiple-choice format and encoded to one-of-K coding, there are 50 variables in the sheet.

Chief complaint text (1,451 variables)

The remote doctor inputs the CC and it contains the primary symptoms identified by the patient. Punctuation marks and numbers in the CC are filtered in data preprocessing to simplify the variables. N-grams ($n = 1, 2, 3, 4$) are extracted from the CC text, and we counted the number of each n-gram where the frequency is greater than six ($= 0.1\%$ of the total record). An n-gram is a contiguous sequence of n items from a given sequence of text or speech. For example, a 2-gram (or bigram) is a two-word sequence, such as “chest pain.” The extracted number of 1-grams (or unigram) was 346, and that of 2-grams, 3-grams (or trigram), and 4-grams was 611, 340, and 154, respectively. The total number of n-grams was 1,451. We placed all n-gram variables in brackets (e.g., back pain) to differentiate the CC variable from the measurement or interview variables.

Training/Testing Set Division and Evaluation

We divided the entire sample into a training set and a testing set at a ratio of 6:4 to evaluate the three prediction models. We constructed each prediction model using the training dataset and evaluated the accuracy of the model by calculating positive predictive value (PPV) and area under the curve (AUC) of the receiver operating characteristic curve (ROC) for each model using the testing set. We repeated this sequence 10 times and calculated the mean and standard deviation of the AUCs, as well as the variable importance.

Analysis methods

In our analysis, we generated three prediction models to determine whether the pharmaceutical for a specific “Drug Function” (“Drugs for”) was prescribed using 1,550 explanatory variables. These three prediction models were generated based on an SVM, Logistic Regression (with Tikhonov regularization), and GBDT. GBDT, as proposed by Friedman [9], produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. Since GBDT is a decision tree model, we also compute the importance of each variable, as well as the predicted probabilities of an input sample as the mean predicted class probabilities of the trees [10]. In this way, we can identify which variables are related to the prescription of each pharmaceutical relative to “Drug Function” and how the variables relate. The probabilities of prescribing a pharmaceutical to the subjects in the testing set were computed to plot the ROC. These probabilities were also used for further clinical decision support, which will be discussed later.

Partial Dependence Plot

A Partial Dependence Plot (PDP) is a visualization method for high-dimensional functions proposed by Friedman [9]. A PDP is useful to visualize the relationship between functions. It plots the marginal effect of the selected variables by integrating out the other variables. Higher value “partial dependence” means higher percentage of ‘yes’ vote by all the weak learners (trees) of GBDT. Let $f(x)$ be the result of a predictor and let x_i be a target explanatory variable. Then, $F_i(x)$ is expressed as follows.

$$F_i(x) = \frac{1}{N} \sum_{j=1}^N f(x_1^j, \dots, x_{i-1}^j, x, x_{i+1}^j, \dots, x_p^j) \quad (1)$$

We can interpret Equation (1) as the predicted mean effects of x_i assuming that all subjects' x_i values are changed to a particular value for x . Since the same dataset, except for the target explanatory variable, is used to calculate $F_i(x_a)$ and $F_i(x_b)$ for x_a and x_b , the difference between $F_i(x_a)$ and $F_i(x_b)$ shows the effect of changing the target explanatory variable from x_a to

x_b . In other words, $F_i(x)$ is a pure outcome against the explanatory variable x_i that removes the effects of other explanatory variables [11]. We use the PDP to interpret the effect of the target explanatory variable.

Results

We built three models using the training set of the data with three different approaches including GBDT, Logistic Regression, and an SVM. We used these three models on the test set of data to predict whether each “Drug Function” is prescribed. With the prediction result, we computed mean AUC of ROCs of predicting each drug function. Then we compared the AUCs of three approaches. GBDT showed the best performance among the three models.

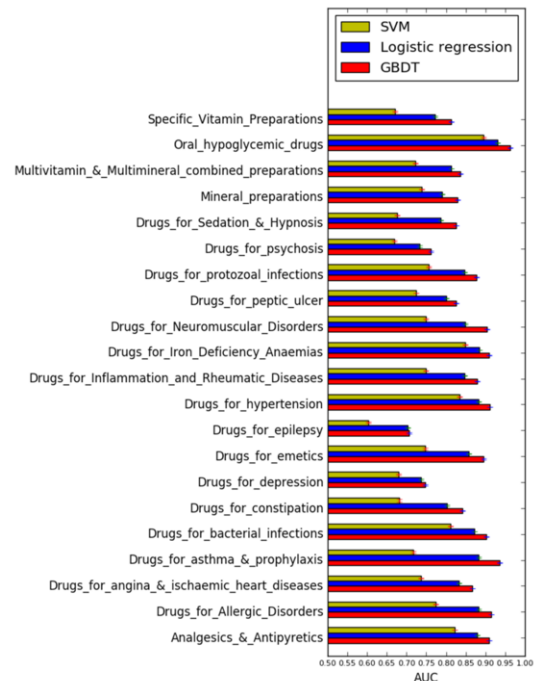


Figure 2- Comparison of AUC of ROC curve of the three methods when predicting the top 21 drug functions

Model Comparison

As shown in Figure 2, GBDT demonstrates better overall performance than the other two models, presumably because a non-linear relationship exists in medical data. The drug function with the highest AUC (Mean = 0.962, Range = [0.944, 0.969]) was Drugs for oral hypoglycemic. Other drug functions with a mean AUC greater than 0.9 were Drugs for Antipyretic Analgesics (Mean = 0.909, Range=[0.899, 0.916]), Drugs for Allergic Disorders (Mean = 0.914, Range = [0.895, 0.932]), Drugs for Asthma & Prophylaxis (Mean = 0.936, Range = [0.924, 0.949]), Drugs for Bacterial Infections (Mean = 0.901, Range = [0.887, 0.916]), Drugs for Hypertension (Mean = 0.916, Range = [0.881, 0.927]), Drugs for Iron Deficiency Anemias (Mean = 0.916, Range = [0.892, 0.929]), and Drugs for Neuromuscular Disorders (Mean = 0.904, Range = [0.887, 0.921]).

Important features

The prediction of drugs for oral hypoglycemic showed the highest AUC (0.962) among 21 most frequent drug function types. The 12 most important features for predicting drugs for

oral hypoglycemic are “bs,” “c_BS” (rule-based triage result of blood sugar based on our diagnosis rules), “uSugar” (urine sugar tape result), “c_usugar” (rule based triage result of urine sugar based on our diagnosis rules), “bmi,” “weight,” “height,” “waist,” “pulse,” “sbp” (systolic blood pressure), “hip” (hip circumference), and “age.”

The relatively high AUC of predicting drugs for oral hypoglycemic is due to the measurement variables and due to feature importance. The 12 most important features are measurement variables or rule-based triage variables generated by measurements.

The 10 times mean of AUC of predicting drugs for peptic ulcers, which is the most prescribed drug type, is 0.826 (95 %CI = [0.812, 0.823]), relatively less than that of drugs for oral hypoglycemic. To better understand this difference, we generated the feature importance for the model for predicting drugs for peptic ulcers. The 12 most important features for predicting drugs for peptic ulcers are “[acidity],” “[pain],” “[pud]” (“peptic ulcer disease”), “[lbp]” (“lower back pain”), “[gastric],” “[abdominal],” “q203: Abdomen” (whether to check abdomen abnormalities as a subjective symptom in an interview), “[gastritis],” “[ache],” “weight,” “[in],” and “bs” (blood sugar). The CC variables are in brackets to differentiate them from measurement and interview variables.

Figure 3 and Figure 4 show the PDP of the top 12 features of these two types of drug functions.

Figure 3- PDP of top 12 features when predicting drugs for oral hypoglycemic

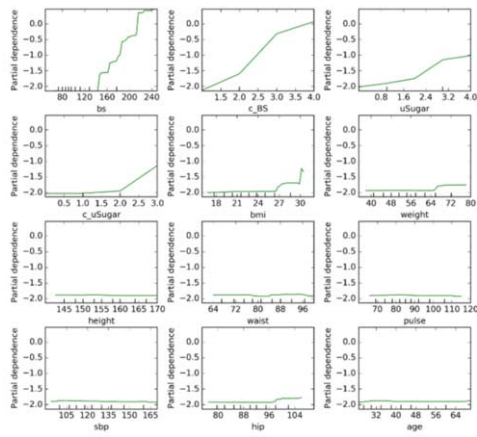


Figure 4 shows that BS greater than 200 mg/dl increases the risk of prescribing drugs for oral hypoglycemic. Similarly, a BMI greater than 30 has the same effect.

The terms with brackets in Figure 3, such as “[acidity],” indicate that the descriptions of these terms in the CC text increases the risk of prescribing drugs for peptic ulcers.

Figure 5 shows a comparison of the mean term frequency in different groups for the CC terms in Figure 4. The vertical axis represents the number of people with the specific term in their CC text, and the green and red bars represent subjects with and without peptic ulcer prescriptions, respectively. The ratios between these two groups are greater than 10 for the terms “acidity,” “pud,” and “gastric,” which also indicates that these terms are highly related to the prescription of drugs for peptic ulcers.

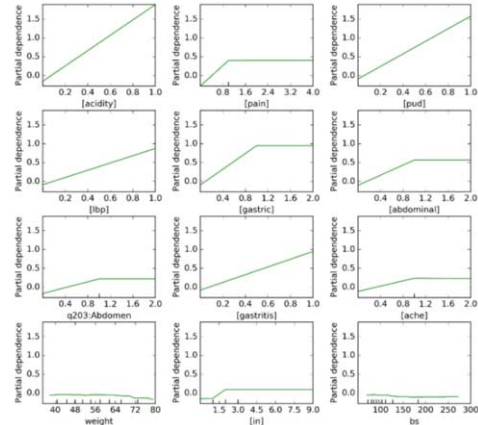
Drugs for hypertension has 669 records and the prediction model for this drug function shows a relatively high AUC (Mean=0.92), similarly with the result of oral hypoglycemic drugs. By computing the feature importance of GBDT model, we found that the variables that affect the prediction signifi-

cantly are diastolic blood pressure, systolic blood pressure, the triage color and age.

Discussion

The prediction model based on GBDT has the highest AUC among the three approaches is supposedly because that logistic regression and SVM (with linear kernel used in this study) are linear models which can separate the data well when the variable of interest (in our case, positive or negative on each drug function) has the linear dependency from the predictors. However, in this study we used predictors such as term frequency of chief complaint text, of which dependency may be different from linear. Whereas, GBDT is able to discover more complex dependencies at the cost of more time for fitting, which may be the reason of GBDT outperforming the other two in this case. The results also suggest that CC variables and measurement variables play different roles in predicting various prescriptions. Predicting drugs for peptic ulcers relies on CC text rather than the measurement of things like BS, body temperature, etc.

Figure 4- Partial dependence plot of top 12 features in predicting drugs for peptic ulcer



In contrast, predicting drugs for oral hypoglycemic relies on measurement variables (particularly BS). We aim to transform the prediction model into a field-deployable application to predict drug functions to support clinical decisions. A prototype program that takes sensor measurements and text information as input and outputs a list of positively predicted prescriptions has been implemented. In addition, a web application is currently under development for practical use.

Figure 6 shows the concept of the diagnosis/prescription prediction web application based on this study. After patient information is input and submitted, the information and output page are shown to a remote doctor. In the output, predictions with a predicted score greater than this subject's score are listed to remind the remote doctor of specific health risks. It is listed by the order of each model's PPV as in Figure 6. PPV is shown in the purpose of providing remote doctor an intuition of the reliability of each drug function prediction based on past data. PPV describes the probability that a patient with a positive test result is actually diseased. For example, if a patient's data is input into the application and hypertension is listed as a health risk, the prediction model computed the score of prescribing drugs for hypertension is 0.7, then PPV of predicting existing data based on this score, 0.7, will be generated and shown as the probability of correctly prescribing for this person basing on past data. This way

provides remote doctor an intuitive understanding on the health risks based on existing data.

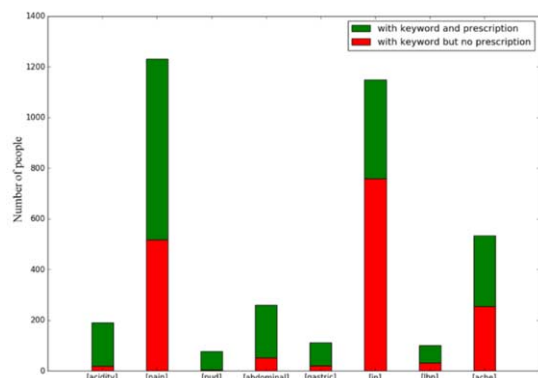


Figure 5- Comparing the mean term frequency in different groups for the term CC in Figure 4

Information Input Example

Sex: ☐ Female ☒ Male Age: 60

Weight: (kg) 80.5 Height: (cm) 173

Systolic blood pressure : (mmHg) 155

Diastolic blood pressure : (mmHg) 90

Blood sugar: mg/dl 170

Blood sugar test type:
☐ Postprandial ☒ Fasting

Chief Complaint Text Input:
 Fever for 1 year. 2)Bodyache for same duration. 3)Acidity

Output Example

Prescription(drug function) with predicted score higher than 0.5	Positive Predictive Value for this subject based on existing data
Oral hypoglycemic drug	93.3%
Hypertension drug	91.1%
Peptic ulcer drug	82.9%

Figure 6- Web application under development for diagnosis/prescription prediction

In this study, we employed a prediction model based on measurement data and CC data. However, for practical use of the prescription prediction application, models with only CC data can be used as symptom checkers for individual users rather than for clinical decision support. The development of a prescription model can be very useful to assist diagnosis and self-diagnosis to tackle the HRH crisis in developing countries.

Conclusion

In this paper, we have compared the results of three machine-learning approaches to predict whether a subject is prescribed pharmaceuticals among the 21 most frequently prescribed drug types. We extracted important factors related to prescribing each pharmaceutical and conducted factor analysis based on tree-based algorithms and natural language pro-

cessing techniques. These important factors and the model can be put to further use to support clinical decisions. Although the goal is to build a reliable system that automatically prescribes drugs for patients, the model is also applicable as a clinical decision support tool, such as identifying potential health risks.

Acknowledgements

The Funding Program supported this research for World-leading Innovative R&D on Science and Technology, "Development of the fastest database engine for the era of very large database and experiment and evaluation of strategic social service enabled by the database engine." The authors appreciate this support.

References

- [1] Karimi-Alavijeh F, Jalili S, Sadeghi M, Predicting metabolic syndrome using decision tree and support vector machine methods, *ARYA Atheroscler*. 12(3):146-152, May. 2016.
- [2] Colubri A, Silver T, Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients, *PLoS Negl Trop Dis*. 18; 10(3). Mar. 2016.
- [3] Mi X, Ikeda H, Nakazawa F, Matsuoka H, Prescription Prediction towards Computer-Assisted Diagnosis for Kampo Medicine, *International Conference on Computer Application Technologies (CCATS)*, pp.126-131, Aug. 2015.
- [4] Soguero-Ruiz C, Mora-Jimenez I, Ivarez JLRA, Hindberg K, Feature selection using Kernel Component Analysis For Early Detection Of Anastomosis Leakage, *Proceedings of 2nd International Workshop on Pattern Recognition for Healthcare Analytics*, Aug. 2014.
- [5] Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Personalized Nutrition by Prediction of Glycemic Responses, *Cell*, 163(5), pp.1079-1094, Nov. 2015.
- [6] Ahmed SM, Hossain MA, Rajachowdhury AM, Bhuiya AU, The health workforce crisis in Bangladesh: shortage, inappropriate skill-mix and inequitable distribution. *Human Resources for Health*, 9(1), 3, 2011.
- [7] Nohara Y, Kai E, Ghosh PP, Islam R, Ahmed A, Kuroda M, Nakashima N. Health checkup and telemedical intervention program for preventive medicine in developing countries: Verification study. *Journal of Medical Internet Research*, 17(1), e2, 2015.
- [8] Bangladesh medicine database, <http://www.bddrugs.com/>
- [9] Friedman JH, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics*, 29(5), pp. 1189-1232, Oct. 2001.
- [10] Natekin A, Knoll A, Gradient boosting machines, a tutorial. *Frontiers in Neuroinformatics*, 7(21), 2013.
- [11] Nohara Y, Wakata Y, Nakashima N, Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation, *Proceedings of the 15th World Congress on Medical and Health Informatics (MedInfo2015)*, p.1073, Aug. 2015.

Corresponding Author

Email address: nnaoki@info.med.kyushu-u.ac.jp

Conditional Density Estimation of Tweet Location: A Feature-Dependent Approach

Hayate ISO, Shoko WAKAMIYA, Eiji Aramaki

Nara Institute of Science and Technology (NAIST), Japan

Abstract

Twitter-based public health surveillance systems have achieved many successes. Underlying this success, much useful information has been associated with tweets such as temporal and spatial information. For fine-grained investigation of disease propagation, this information is attributed a more important role. Unlike temporal information that is always available, spatial information is less available because of privacy concerns. To extend the availability of spatial information, many geographic identification systems have been developed. However, almost no origin of the user location can be identified, even if a human reads the tweet contents. This study estimates the geographic origin of tweets with reliability using a density estimation approach. Our method reveals how the model interprets the origin of user location according to the spread of estimated density.

Keywords:

Social Media; Geographic Mapping; Disease Outbreak

Introduction

The recent rise in popularity and scale of social media has created a growing necessity for social-media-based public health surveillance. The feasibility of such approaches has been demonstrated using various associated information, including temporal information [1-2] and spatial information [3-5]. Temporal information is associated with all tweets, but spatial information is often unavailable for privacy reasons. One report described that fewer than 0.5% of tweets include GPS information [6]. This problem has become an important motivation for many studies of location estimation [7-11].

Recent studies have elucidated the characteristics of geo-tagged tweets using various approaches. These include location specificity of user attributes such as gender and age [12], linguistic variation [13], temporal effects on location classification accuracy [14], population biases [15], and content-based geographic density of tweets [16]. Our study is aimed at exploring the content-based characteristics of tweet location (System estimated location vs. Geo-tagged location) further. We also investigate differences between the estimated results and the interpretability of human estimation (System estimated location vs. human estimated location). The motivating examples of comparison to humans are shown in Figure 1. Based on these examples, even a human would have difficulty estimating the precise location. However, by some clues, we were able to infer the tweet origin weakly. As this example shows, we can have an idea of a tweet's general region of origin. Therefore, unlike previous studies aimed at estimating the concrete region, our task is to estimate the probability density of the origin location, which more naturally fits human understanding.

A recent study [16] was undertaken to estimate the location as a density estimation problem. Although their motivations resemble ours, this research represents the geo-location

identifiability of a given tweet as a combination of word-specific or n -gram-specific Gaussian Mixture Model (GMM).

We summarize the contributions of this study as follows:

- We provide simpler and more reasonable approaches to estimate the geographic region of a tweet. Although an earlier study [16] estimated GMM in each word independently, our method handles tweet contents in a vector representation.
- We examine the relation between human inference and geographic biases of geo-tagged tweets.
- We objectively and quantitatively evaluate the differences between human and model inferences.

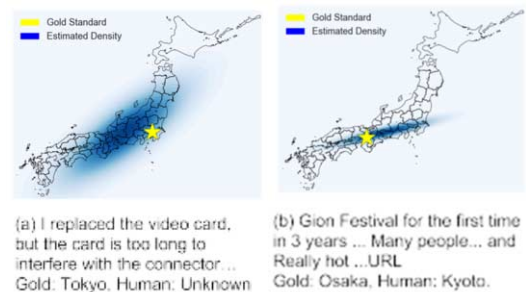


Figure 1—Estimated geographic distribution of a tweet. “Gold” represents the true origin of the tweet and “Human” represents the human interpretation of the origin of user location from the tweet content.

Materials

Tweet Dataset

We collected 554,320 geo-tagged Japanese tweets for a week (July 15, 2012 – July 21, 2012). For our purposes, we extracted tweets that were posted by Official Twitter clients: *Twitter for iPhone* and *Twitter for Android*. Consequently, our corpus comprises 204,748 tweets.

To estimate the number of mixture components, we split tweets into a triplet of training, validation, and test data. We used 144,748 tweets for training data and 30,000 tweets each for validation and test data. This is the section where the authors describe the methods used at the level of detail necessary to convey the sample size, setting, procedure, datasets, analytic plan, and other relevant particulars to the reader.

Human Annotation Rule

To evaluate how our model interprets origin locations of tweets, humans annotated Japanese region (8-way) or prefecture (47-way) labels for 5,000 tweets that were sampled randomly from the test data. Tweet data with several granularity labels were annotated independently by two annotators. To avoid birthplace and residential place biases, we permitted rough searching for

the word included in a tweet to identify its origin location. The human agreement rate was 93.6%, as measured by 100 tweets sampled randomly from the annotated ones. Table 1 presents the results, which indicate the difficulty of this task, in which most tweets were not identified by humans. Furthermore, goodness-of-fit test results showed that tweets with a position were identified by human beings. Results also showed that the multinomial distributions by prefecture differ for learning data tweets.

Table 1-Human annotated data summary: goodness of fit (GoF) test. We tested humans' correctly annotated data with overall geo-tagged tweets as a multinomial distribution at the prefecture level.

Level	Ratio (%)	Precision (%)
Annotated	10.4	87.8
Unknown	89.6	-
Prefecture	9.1	89.7
Region	1.3	74.6

Methods

Word-Specific Gaussian Mixture Model

In an existing approach [16], the GMM has been used for geographic density estimation of geo-tagged tweets. Each tweet is converted to n -gram features consisting of the number of n -gram occurrences in a corpus and geographic coordinates (longitude and latitude) of n -grams. GMM applied for each n -gram w_j is defined as word-specific GMM as

$$1) p(y|w_j) = \sum_{k=1}^K \pi_{kj} N(y|\mu_{kj}, \Sigma_{kj})$$

here $y \in \mathbb{R}^2$ represents geographic coordinates (latitude and longitude), w_j represents the word indexed in j , π_{kj} is the weight as the word w_j is assigned to the k -th mixture component, and $N(y|\mu_{kj}, \Sigma_{kj})$ is a multi variate Gaussian distribution with mean μ_{kj} and covariance Σ_{kj} .

After estimating GMM for each n -gram, the weighted sum of word-specific GMM is combined as

$$2) p(y|x) = \sum_{j=1}^J \pi_j p(y|w_j)$$

Where $x = \{w_1 \dots w_J\}$ represents the words in a tweet, π_j is the weight of the GMM on the word w_j .

For the formula, it is important to ascertain the weight of each GMM $\pi_j(w_j)$. To date, most methods calibrate parameters for improving the prediction accuracy, but they merely consider the geo-location identifiability of n -grams to adjust weights. These approaches also merely consider whether a specific word is included or not. They do not consider the meanings of the respective tweets.

In contrast, this study applies *Gaussian Mixture Regression*, which allows expansion of word-specific GMM further by ascertaining weights of tweet GMM automatically from joint probability distributions of a tweet and geo-location. Consequently, we can impose any kind of feature to our model.

Feature-Dependent Density Estimation

To represent our location density model, we use Gaussian Mixture Regression (GMR) [18], which is formalized by a conditional distribution of a jointly estimated Gaussian Mixture Model (GMM).

Our model need not prepare a specific evaluation index for feature-dependent weight estimation. We can derive the weight of two-dimensional GMM from jointly estimated GMM. Therefore, we designate our model as *feature-dependent*. Depending on given feature vectors, the feature of GMR varies the mixture of weights and each component of Gaussian location and variances. Figure 1 presents examples of conditional geographic density of two tweets.

To obtain the GMR results, we first estimate the joint probability $p(y,x)$ of p -dimensional tweet representation and two-dimensional Gaussian Mixture Model [17]. Then conditional distribution $p(y|x)$ of geo-location y for a given tweet x can be derived analytically from the joint distribution $p(x,y)$ as follows:

$$3) p(y|x) = \frac{p(x,y)}{\int_x p(x,y)dx} = \frac{p(x,y)}{p(x)} = \frac{\sum_{k=1}^K \pi_k N(x,y; \mu_k, \Sigma_k)}{\sum_{k=1}^K \pi_k N(x; \mu_k, \Sigma_{k,xx})}$$

where

$$4) \mu_k = \begin{pmatrix} \mu_{k,1} \\ \mu_{k,2} \end{pmatrix}, \Sigma_k = \begin{pmatrix} \sigma_{k,1}^2 & \rho_k \sigma_{k,1} \sigma_{k,2} \\ \rho_k \sigma_{k,1} \sigma_{k,2} & \sigma_{k,2}^2 \end{pmatrix}$$

$$5) N(y, \mu_{k,y|x}, \Sigma_{k,y|x}) = \frac{N(x,y; \mu_k, \Sigma_k)}{N(x; \mu_k, \Sigma_{k,xx})}$$

$$6) \mu_{k,y|x} = \mu_{k,y} + \Sigma_{k,yx} \Sigma_{k,xx}^{-1} (x - \mu_x)$$

$$7) \Sigma_{k,y|x} = \Sigma_{k,yy} - \Sigma_{k,yx} \Sigma_{k,xx}^{-1} \Sigma_{k,xx}$$

$$8) \pi_k = \frac{\pi_k N(x; \mu_k, \Sigma_{k,xx})}{\sum_{k=1}^K \pi_k N(x; \mu_k, \Sigma_{k,xx})}$$

The key point of GMR is that the weights of mixture parameters are changed flexibly depending on the feature vector x . Consequently, the tweet's vector representation defines the two-dimensional geographical probability density functions $p(y|x)$. GMR is useful for any feature as p -dimensional vectors.

As described in the paper, we use a continuous word vector learned by *fasttext* [19] to compress the dimensions of our tweet dataset. We compose a vector representation of tweet by averaging all the word vectors in tweets.

Results

To compare our model with human inference, we evaluated our model through several perspectives.

First, we calculated the distance between our model density and the origin of tweets in a different identifiability dataset such as the prefecture level or region level. We choose the mode value of the estimated distribution as the estimated location and we got the city name from estimated geographic coordinates using Google Map API.

As a baseline method, we used the regularized linear regression method, Elastic-Net [20]. We optimized the baseline model using the validation set.

Our first results are presented in Table 2 and 3. Although our model performs worse than the baseline model, our model monotonically improves prediction performance through human inference improvement.

Second, we comprehensively investigated cases in which our model revealed a result similar to a human's inference, when it failed to estimate origin locations, and when it outperforms human inference. Our characteristic examples are presented in

Figure 2. As the first example, Figure 2(a), shows the human label as coincident with the origin of tweets, in such cases, our model easily estimates the origin of locations. In the second example, our model has failed estimation with high accuracy. Our model density is widely spread. However, humans are also puzzled when specifying the location and labeled the region level. The last example reveals the superiority of density-based estimation. In the tweet, a user stated “Yokohama”, which is a city of Kanagawa prefecture. Therefore, the human annotated Kanagawa prefecture for this tweet.

However, the speaker makes remarks at the boundary of the prefecture; the speech position is actually within Chiba prefecture. In a conventional classification approach, we misclassified this tweet, but it is apparent that our method estimates the distribution across both prefectures.

Our model incorporates the uncertainty of the user location estimated from tweet contents. We ascertained that a density-based approach is more reasonable to cover a wider range of class such as a prefecture or region than classification approaches for a difficult tweet to identify the geolocation. To improve the model validity further, feature vectors have an important responsibility. Although this research only employs the textual information for geolocation estimation, the many previous researches empower the geolocation performance via classification. We will further explore which feature has good effects for estimating the tweet’s uncertainty.

In addition, we will apply our model for non-geotagged infectious diseases related tweets such as Influenza [2] to explore the regional trends of the infectious diseases.

Conclusion

In this study, we demonstrated that GMR provides a new perspective for estimation of the tweet posting origin. We provide simpler and more reasonable approaches to estimate the geographic region of a tweet. Although an earlier study [16] estimated GMM in each word independently, our method handles tweet contents in a vector representation. We examine the relation between human interpretability and geographic biases of geo-tagged tweets.

Table 2-Mean distances: Region data include Prefecture data

Level	GMR (km)	Elastic-Net (km)
Prefecture	251	271
Region	242	268
Overall	278	272

Table 3-Median distances: Region data include Prefecture data

Level	GMR (km)	Elastic-Net (km)
Prefecture	154	181
Region	134	181
Overall	214	191

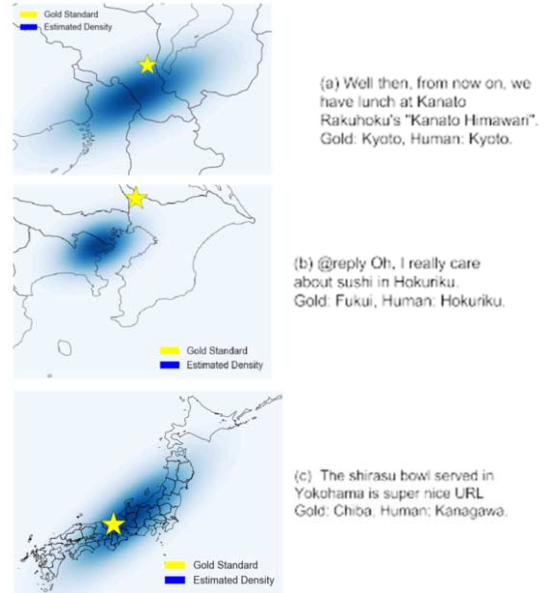


Figure 2-Characteristic estimated density examples.

Acknowledgements

Funding sources for the work and other relevant acknowledgements are noted here. Authors may also present disclosures or disclaimers to their work in this section.

References

- [1] M.J. Paul, M. Dredze, and D. Broniatowski, Twitter improves influenza forecasting, *PLOS Currents Outbreaks* (2014).
- [2] H. Iso, S. Wakamiya, and E. Aramaki, Forecasting word model: Twitter-based influenza surveillance and prediction, in: *Proceedings of the 26th International Conference on Computational Linguistics*, 2016, pp. 76-86.
- [3] D.A. Broniatowski, M.J. Paul, and M. Dredze, National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic, *PLoS One* 8 (2013), e83672.
- [4] A. Sadilek, H. Kautz, and V. Silenzio, Predicting disease transmission from geo-tagged micro-blog data, in: *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
- [5] S. Eubank, H. Guclu, V.A. Kumar, M.V. Marathe, A. Srinivasan, Z. Toroczka, and N. Wang, Modelling disease outbreaks in realistic urban social networks, *Nature* 429 (2004), 180-184.
- [6] Z. Cheng, J. Caverlee, and K. Lee, You are where you tweet: a content-based approach to geo-locating twitter users, in: *Proceedings of the 19th ACM international conference on Information and knowledge management*, ACM, 2010, pp. 759-768.
- [7] O. Ajao, J. Hong, and W. Liu, A survey of location inference techniques on Twitter, *Journal of Information Science* 41 (2015), 855-864.
- [8] D. Jurgens, T. Finethy, J. McCorriston, Y.T. Xu, and D. Ruths, Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice, in: *ICWSM*, 2015, pp. 188-197.
- [9] B. Han, P. Cook, and T. Baldwin, Text-based twitter user geolocation prediction, *Journal of Artificial Intelligence Research* 49 (2014), 451-500.
- [10] A. Rahimi, D. Vu, T. Cohn, and T. Baldwin, Exploiting text and network context for geolocation of social media users, *arXiv preprint arXiv:1506.04803* (2015).
- [11] Z. Liu and Y. Huang, Where are You Tweeting?: A Context and User Movement Based Approach, in: *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, ACM, 2016, pp. 1949-1952.
- [12] U. Pavalanathan and J. Eisenstein, Confounds and consequences in geotagged Twitter data, *arXiv preprint arXiv:1506.02275* (2015).
- [13] J. Eisenstein, B. O'Connor, N.A. Smith, and E.P. Xing, A latent variable model for geographic lexical variation, in: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2010, pp. 1277-1287.

- [14] M. Dredze, M. Osborne, and P. Kambadur, Geolocation for twitter: Timing matters, in: *Proceedings of NAACL-HLT*, 2016, pp. 1064-1069.
- [15] M.M. Malik, H. Lamba, C. Nakos, and J. Pfeffer, Population bias in geotagged tweets, in: *ICWSM Workshop on Standards and Practices in Large-Scale Social Media Research*, 2015, pp. 18-27.
- [16] R. Friedhorsky, A. Culotta, and S.Y. Del Valle, Inferring the origin locations of tweets with quantitative confidence, in: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, ACM, 2014, pp. 1523-1536.
- [17] T.S. Ferguson, A Bayesian analysis of some nonparametric problems, *The annals of statistics* (1973), 209-230.
- [18] H.G. Sung, *Gaussian mixture regression and classification*, Rice University, 2004.
- [19] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, Bag of tricks for efficient text classification, *arXiv preprint arXiv:1607.01759* (2016).
- [20] H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (2005), 301-320.

Address for Correspondence

Eiji ARAMAKI, Ph.D. <aramaki@is.naist.jp>

Analysis of the Existence of Patient Care Team Using Social Network Methods in Physician Communities from Healthcare Insurance Companies

Márcia Ito^a, Ana Paula Appel^a, Vagner Figueredo de Santana^a, Luis G. Moyano^b

^a IBM Research, Sao Paulo, SP, Brazil,

^b Department of Statistical Physics and Complex Systems, CONICET and Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Mendoza, Mendoza, Argentina

Abstract

Care teams are formed by physicians of different specialties who take care of the same patient. Hence, if we find physicians that share patients with each other probably they configure an informal care team. Thus, the objective of this work is to explore the possibility of finding care teams using Social Network Analysis techniques in physician-physician networks where the physicians have patients in common. For this, we used healthcare insurance claims to build the network. There was the agreement on the metrics of degree and eigenvalue and of betweenness and closeness, also physicians with the 5 highest eigenvalues are highly interconnected. We discuss that the analysis of the physician-physician network with metrics of centrality is promising to reveal informal care teams. The high potential in calculating these metrics is verified from the results to evaluate member's performance and with that how to take actions to improve the work of the team.

Keywords:

Data Mining; Patient Care Team; Quality of Life

Introduction

The control and management of chronic diseases are one of the greatest challenges of health and economic development in the 21st century as they are directly related to the population quality of life as well as to world productivity and spending. It is estimated that annually approximately 16 million people die prematurely before the age of 70 due to chronic diseases. There is also a loss of US \$ 7 trillion in health productivity and spending over the next 15 years if no action is taken [1]. Chronic patients usually have the main diagnosis and other associated diseases requiring care by more than one medical specialty as well as other professionals who guide them in changing their behaviors as nutritionists and physiotherapists, for example. Professionals who care for the same patient form a team called care team. Coordinating the care activities of these professionals is essential to optimize the use of health resources, called care coordination. In addition, according to Owens[2], the approximate expenditures of a patient in whom there is no coordination of care is 75% higher than those who count on this team. The author also suggests that by improving care coordination, patient costs can be reduced by 35%. The concept of patient health management through coordinated care teams has been explored by several authors since the late 1960s. Care coordination aims to reduce the fragmentation of care and improve delivery of health services. Thus, there are several efforts to implement this type of program [3]. Studies[4],[5] indicate that successful programs are those in which long-term, trusting relationships are built between the patient and the care

team and among the members of the care team. Collaboration among members of a care team is the most important feature as it allows: (1) planning the treatment in a shared way, (2) facilitating decision making, (3) defining the objectives of patient care, (4) coordinating in a transparent way in which each member assumes responsibility for his / her performance with the patient, (5) working cooperatively, and (6) facilitating communication among the members [6],[7],[8].

Since 1995 and with a considerable increase in the last years we found research that uses the techniques of SNA (Social Network Analysis) to evaluate the impact of networks of health professionals in hospitals [9],[10],[11],[12],[13],[14]. At the outset, these studies aimed to verify if it was possible to extract the network of physicians through the existing data in hospitals and to analyze improvements in the quality and safety of the patient [9]. More recently, there are studies that look for metrics and indicators in these networks that relate the network of physicians to costs and/or re-hospitalizations in hospitals [10],[11],[13],[14]. The conclusion from these papers is that the more that physicians collaborate with each other the costs tend to be lower. In this case, they used the calculation of density, degree, and betweenness. Readmission in hospitals can be seen by the density in which hospitals with a high number of medical communities have a low readmission of their patients [10]. Another study presents how to find medical errors and frauds in health insurance companies [12].

However, we haven't found any work using SNA techniques that focuses on the relationship between physicians who have common patients who are not hospitalized. Thus, this study aims to explore whether it is possible to find care teams in networks of physicians from health insurance company claims and the possibility of using centrality metrics to analyze the interaction between these informal care team members.

Materials and Methods

Source Data

Health insurance companies generate a large number of data as products of their daily operations. Operations in this area produce a considerable amount of data, specialty, transactional, and also other forms of data, such as demographics, service provider location, and other essential data to accurately carry forward all business processes. These data are related to their demands, such as the claims that are the information that generates the payments to be made to the providers of health services. The claims inform all the details about the procedures and consultations carried out by the providers, such as the professional who carried out the procedure, the procedure itself, and the insured who received the procedure among others.

In this paper the definitions used, from this moment on, are:

- **Insured:** represents in the database provided, each of the individuals who used health insurance services and who at times can be referenced as a **patient**;
- **Physician:** represents each of the physicians who provided services or are registered as such in the source database;
- **Provider:** represents healthcare institutions that provide services to insurances and in which physicians assist the insured.

The data analyzed in this study correspond to the claims originated in operational processes of a large Brazilian health insurance company. The 18-month claims that cover the period from January 2013 to June 2014 were provided. The data set includes medical procedures performed throughout Brazil and all claims paid by the insurer and corresponding to a total of 108,982,593 claims, this means that in average the health insurance company processes almost 300,000 claims per day. The rationale of using this data resides not only because of its volume but also because it represents the main and more complex operation of the health insurance company. Permission to use the data for the purpose of the research was made through a contract between both parts (research department and the health insurance company). Before the anonymization of physicians and patients, the data was cleared (withdrawal of invalid registrations in Regional Medicine Council). In the original data we had information about 279,085 physicians, being 81% valid (related with 73% of claims); by valid we mean that the physician register was well formed and the number informed to the provider was valid. Moreover, we had information about 2,243,198 patients and 26,033 providers.

A common situation in Brazil about physicians and service providers is that some providers use one or a few physician IDs to register all the claims to the health insurance company. This fact occurs when only one physician is registered in the insurance company and can distort the data interpretation, affecting the quality of the results. Thus, in order to identify provider-physician relationships that are anomalous, we selected providers that are strongly related to some physicians. We consider providers with 90% of their relationship with one physician that has a large number of claims. In this selection, we removed 500 physicians (about 0.22% from total) from our data in order to prevent any misleading in the interpretation. After this selection, we had 12,924 providers connected with 219,675 physicians, that pair provider-physician which were analyzed in this project. Still related to the data, two important aspects are related to data distribution considering states and physician specialties. First, about country states, we have around 25% of the data without state information. This proportion becomes bigger when we look at the data as a graph since to consider the relationship between physicians the missing value of state in one or both nodes (physicians) invalidate an important piece of our analysis, i.e., the connections between physicians.

Finally, the specialty is another important information to explain the relationship between the physicians, because when we find a relationship between them of related specialties, we can conclude that they form an informal care team of the patients in common. However, because of the large amount of missing values and no specified specialty, the use of this information was limited. In the database table with information about physicians, we had 658,543 tuples that belong to the physicians that work with the health insurance company. Among these, 436,981 (66%) are active physicians (this means that they are active in the company's database and their

registries are valid in the Regional Medicine Council) and only 29.3% have the specialty information.

Mapping Claim Data as a Graph

Health insurance claims data may be mapped in graph form by considering the healthcare professional as a node (of a certain type), patients also as nodes (of a different type), and establishing a link whenever there is a claim mentioning both entities (Figure 1). The resulting graph is an example of a bipartite graph, where there are several types of nodes and there are no links between nodes of the same type. Moreover, additional bits of information in the data may be included in the graph in the form of weights in the links (such as the timestamp, the service provided or the expense amount), or as attributes in the nodes (as the patient demographics information or the medical specialty of the healthcare professional). Both weights and attributes may be represented in a graph layout by mapping their values to colors or sizes in the case of nodes, or colors and widths in the case of links. In this work, we focus on the relative influence of the physicians' activity in the relationship among the physicians in the health insurance company network. Two physicians are considered related if they have a common patient (a patient that had a consultation with both physicians). This does not indicate a direct relationship between the physicians but if they have a very large number of common patients (represented as an outlier) there is a high probability that physicians have some kind of professional relationship, for example, they have a similar profile, same provider, work nearby, a similar education background, etc. For this, we build a graph $G = (V, E)$ physician-physician, where the $|V| = N$ denote the set of nodes that are physicians and $|E| = M$ denotes the edges, which means physicians that have consultation claims with patients in common, $e_k \in E$ e $e_k = \{(v_i, v_j) | v_i, v_j \in V\}$.

The fact that physicians attend to the same patient does not necessarily mean that there is collaboration between them in the treatment of the patient, but if it is very recurrent it may indicate this possibility. On the other hand, if the number of patients common among them is high there is the probability that they have some kind of professional relationship, for example, to be the team care of those patients.

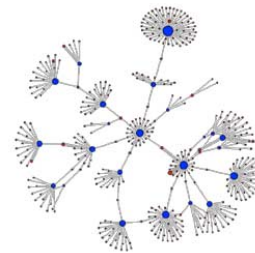


Figure 1 – Example of the patient-physician network, which can be used to map the relationship between physicians

Results

First of all, centrality metrics were selected to analyze the relative importance of physicians in the network and if there are any possible collaborations among them in the treatment of patients:

- **Degree:** The importance of the physician is proportional to the number of patients shared with other physicians.

- **Eigenvalue:** The higher the number of shared patients with other important physicians, the more important the physician is considered. If the physician shares a large number of patients but with physicians not considered important in the network, the physician is not considered important [15].
- **Betweenness:** Taking into account that physicians can be influential to other physicians that are closer, for example, when a physician indicates another physician to his/her patient, it is possible to define a metric that captures this proximity [16].
- **Closeness:** The physicians can be related by the amount of other physicians that are known among each other, indirectly, without their knowledge [9]. Thus, it is possible to compute the degree of separation between physicians by considering the number of physicians that are needed to connect two physicians.

We used all these four metrics to the same network including all the states, considering also how they change over time. We divided the whole period into intervals, considering quartiles. In Table 1, we can see that the number of physicians and patients that we use in our analyses. The results allow associating of the metrics to all physicians connected to a main group (major related component) within the quartiles time interval (Q2 to Q4 of 2013 and Q1 to Q2 of 2014).

Table 1 – Number of physicians and patients analyzed by quartile

Quartile	Q2 2013	Q3 2013	Q4 2013	Q1 2014	Q2 2014
Physician	8,443	8,668	7,870	8,273	8,273
Patient	215,252	225,708	197,018	213,635	203,116

The resulting ranking depends on the used metric, but there is still a degree of correlation among the metrics. For instance, for Q2 of 2014, 14 physicians from top 20 using **eigenvalue** as a metric, are the same in the others metrics, showing us that the four centrality measures are consistent. Given that all the metrics capture the importance of the physicians in a similar way. In general, we have a metric concordance between 10%

Table 2 – Temporal evolution by quartile of top 10 physicians with highest centrality measures in the physician-physician network, considering all states.

Physician	Specialty	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2
M _{sp} 153	Cardiologist	1	1	2	1	1
M _{sp} 154	Endocrinologist	2	3	1	2	2
M _{sp} 164	Dermatologist	3	4	7	4	5
M _{sp} 142	Hematologist	4	5	3	3	3
M _{sp} 242	Nephrologist	5	10	10	14	12
M _{sp} 243	Orthopedist	6	8	8	6	8
M _{sp} 140	Otorhinolaryngologist	7	2	4	5	4
M _{sp} 148	Cardiologist	8	7	5	13	11
M _{sp} 149	Hematologist	9	12	14	12	16
M _{sp} 244	Not declared	10	13	12	15	10

We noticed that there is a stable behavior in the position of the physicians over time, for example, the physician M_{sp}153 holds the first position in all quartiles except the fourth quartile of 2013 and remains in the second position (not getting far from the previous position). In this way, it is possible to measure

and 20% of top 100 physicians, which is high considering that we have almost 8k physicians.

Figure 2 shows the degree distribution of the physician-physician network. As we can see, the decay is much smaller than an exponential curve which is the case of a random graph and a little more curve than a straight line that is a perfect power law. So, it shows that this behavior is similar to a graph of a social network than to a synthetic graph, meaning that the behavior of physicians is not random but strongly dominated by correlations such as patients returning to the doctor, similar specialties, proximity to providers, etc. [17] This supports our approach in using SNA methods to analyze the presented data.

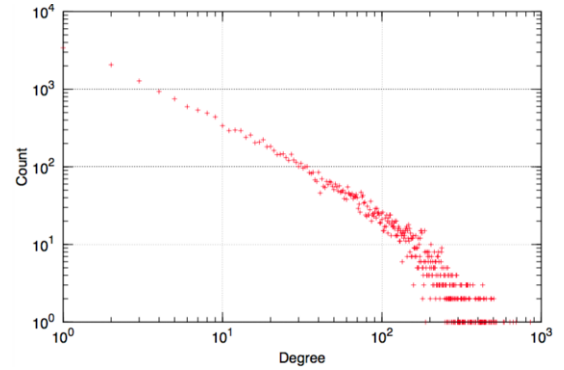


Figure 2 - Degree distribution of the physician-physician Network

One important characteristic is the possibility of following the physician's evolution related with these metrics. In Table 2 we show the top 10 physicians over the period and considering all states using eigenvalue. As we can see some physicians have a stable behavior over time, meaning they are occupying closer positions, for example, physician M_{sp} 153 is kept in the first position during most of the time except in Q4 of 2013 when he/she moved to the second position. In other cases, a physician has an ascendant or descendant movement, which can indicate a change in the physician area or patients' behavior that might be interesting for the health insurance company.

these changes and thus know how much he/she is collaborating/influencing other physicians.

Another result obtained is that even having a high number of physicians (around 8k per quartile), the top 5 with the highest value in the eigenvalue metric in the second quartile of 2014,

besides their higher connections, they are also highly connected to each other, meaning that each one share patients with the other 5. When checking the specialty of each one we have an otorhinolaryngologist, a dermatologist, an endocrinologist, a hematologist, and a cardiologist. This fact, when analyzed in relation to the specialties that are connected, strongly suggest a collaboration between them in the care of the patient.

Discussion

According to Abdelzaher [13], patient sharing among physicians increases the likelihood of interaction and the sharing of information between them, as well as a chance to care for their patients in a synchronized way. Then, it can be concluded that the sharing of the patient between physicians strongly indicates that they form a care team. On the other hand, studies by Uddin [10] concluded that hospital physicians with high value in degree metrics indicate that they collaborate more with peers. Therefore, it is possible to conclude that physicians with high value in degree metrics collaborate and share information among each other, as well as interaction and synchronization of care to patients who are common to them. This all characterizes care coordination and thus they form a care team. Following the same reasoning, it is possible that physicians high values in degree metrics and eigenvalue, even not belonging to a hospital, possibly form a care team with their patients in common. This can be verified when analyzing the top 5 physicians who we studied, that demonstrate they are highly connected. According to Stukel et al. [18], providers tend to form informal care teams with physicians of different specialties only based on their share of patients and information. The high connectivity among the 5 top physicians suggests a sharing of patients and information and thus a formation of an informal care team.

In addition, his/her specialties (dermatologist, endocrinologist, cardiologist, otorhinolaryngologist, and hematologist) suggest the formation of a care team for chronic patients, since according to Stukel et al. [18], the chronic disease care is better when we have large multispecialty physician group practices. Such practices reduce complications and increase hospital and emergency readmission.

Beyond that, studies reported in Cunningham [9] assert that when we have cohesive and collaborative networks (of professionals or health institutions) it is beneficial to patient care and safety, so these informal care teams indicate that patient care and safety are probably good and should be supported and encouraged by providers.

The high-value metric indicates which physician has the most importance in this team, so the absence of him/her can damage the other relationships or even dissolve the care team. When analyzing the top 5 physicians the fact seems to make sense, because the most important physician is the cardiologist and the second is the endocrinologist. These specialties are usually present in the treatment of chronic diseases and thus this fact corroborates the analysis performed and the case studied.

From the Uddin's research [10], doctors who have many connections (high value in the degree metric) suggest that they have more knowledge on how to assist their patients, since it makes it easier for the physician to share knowledge effectively with his or her other peers. In addition, it was found that the costs of hospitalization and readmission of patients are lower when the treatment is performed by highly connected physicians. Thus the physician with high values in the degree metric indicates that he/she belongs to several care teams and that he/she has a great amount of knowledge to care his/her patients, meaning better quality in the assistance and optimization in the use of resources.

Regarding the metric of betweenness, we can conclude that the greater the fluctuation is, the collaboration is less and the communication is impaired among the physicians [10]. Thus, we can define that finding this type of fluctuation between the members of a care team, probably, means that the treatment did not happen in a synchronized way and therefore is a candidate to have a higher health cost to the provider.

Finally, it is possible to analyze the relationship evolution of the physicians in the teams, when in this study we find the possibility to analyze the relationship of the medical community by quartile. This is important because in this way the provider can analyze the care teams and make decisions regarding the need for training, counseling, and meetings with all the stakeholders involved in order to improve the quality of the care service.

Conclusion

In this work it was possible to analyze the degree, eigenvalue and betweenness metrics that were very promising to define probable informal care teams by finding subsets of doctors with a high index of common patients. The high potential in calculating these metrics is verified in the members of the care team and from their results and from evaluating their performance. Another important point is to be able to evaluate over time the relationship evolution of these doctors in the team and with that take actions to improve the work of the team.

As future work, a more detailed analysis will be performed of a group of physicians highly connected in the graph ("informal care teams") and not highly connected. Analysis of what degree of connectivity is needed to know whether or not they actually form an informal care team will be carried out. Analysis of whether the cost of care of patients with the same diagnoses is lower than those with the same types of patients being taken care by physicians who are not highly connected will also be done. We can also analyze the impact the cost of patients being taken care by the set of doctors who have fluctuation in the betweenness (low communication and interaction among doctors) to see if there is any impact on the cost of treatment.

References

- [1] World Health Organization. *Noncommunicable diseases progress monitor*. Switzerland, WHO, WHO Press, 2015.
- [2] M. K. Owens, Costs of uncoordinated care. In: Yong PL, Saunders RS, Olsen, LA, editors. *The healthcare imperative: Lowering costs and improving outcomes. Workshop series summary. Washington, DC, National Academy Press*, (2010),109-140.
- [3] K. M. McDonald, V. Sundaram, D.M. Bravata, R. Lewis, N. Lin, S. Kraft, et. al. Care coordination. Vol 7. of: Shojania KG, McDonald KM, Wachter RM, Owens DK, editors. *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies, Technical Review 9 (Prepared by the Stanford University-UCSF Evidence-based Practice Center under contract 290-02-0017)*, AHRQ Publication no. 04(07)-0051-7, Rockville, MD, Agency for Healthcare Research and Quality, 2007.
- [4] S. Van Houdt, W. Sermeus, K. Vanhaecht, J. De Lepeleire. Focus groups to explore healthcare professionals' experiences of care coordination: towards a theoretical framework for the study of care coordination. *BMC Family Practice* 15 (2014), 177.
- [5] E.H. Wagner, N. Sandhu, K. Coleman, K. E. Phillips, J. R. Sugarman. Improving care coordination in primary care. *Med Care* 52(11 Suppl 4)(2014),S33–S38.
- [6] S.J. Cunningham, S. M. Dillon. Authorship patterns in information system. *Scientometrics* 39 (1997), 19-27.
- [7] T. Luukkonen, R. J. Tijssen, O. Persson, G. Sivertsen . The measurement of international scientific collaboration. *Scientometrics* 28 (1993), 15-36.
- [8] D. B. Beaver, R. Rosen. Studies in scientific collaboration. *Scientometrics* 1 (1978):1, 65-84. apud ¹⁰

- [9] F. C. Cunningham, G. Ranmuthugala, J. Plumb, A. Georgiou, J. I. Westbrook, J. Braithwaite. Health professional networks as a vector for improving healthcare quality and safety: a systematic review. *BMJ Qual Saf* 21 (2012), 239-249.
- [10] S. Uddin, L. Hossain. Effects of physician collaboration network on hospital outcomes. In: *Proceedings of the fifth Australian Workshop on Health Informatics and Knowledge Management (HIKM 2012)*. Melbourne, Australia, 2012, 67-73.
- [11] C. E. Pollack, G. E. Weissman, K. W. Lemke, P. S. Hussey, J. P. Weiner. Patient Sharing Among physician and cost of care: A network analytic approach of care coordination using claims data. *J Gen Intern Med* 28:3 (2012), 459-65.
- [12] V. Chandola, S. R. Sukumar, J. Schryver. Knowledge discovery from massive healthcare claims data. In: *Proceedings of KDD'13*, Association of Computing Machinery (ACM), Chicago, USA, 2013, 1312-1320.
- [13] A. F. Abdelzaher, P. Ghosh, A. F. Mussawi, D. S. Jackson, J. A. Palesis, J. P. Deshazo. Care coordination metrics of patient sharing among physicians: A social network analytic approach. In: *Proceedings of BCB'14*, Association of Computing Machinery (ACM), Newport Beach, USA, 2014, 635-636.
- [14] S. Uddin, M. Kelaher, M. Piraveenan. Impact of physician community structure on healthcare outcomes. *Stud Health Technol Inform*. 214 (2015), 152-8.
- [15] B. Ruhnau. Eigenvector centrality a node centrality? *Social networks*, 22:4(2000),357-365.
- [16] M. Barthélemy. Betweenness centrality in large complex networks. *The European Physical Journal B -Condensed Matter and Complex Systems* 38:2 (2004), 163-168.
- [17] M. Newman. *Networks: an introduction*. OUP Oxford, 2010.
- [18] T. A. Stukel, R. H. Glazier, S. E. Schultz, J. Guan, B. M. Zagorski, P. Gozdy, D. A. Henry. Multispecialty physician network in Ontario. *Open Medicine* 7:2 (2013), 40-55.

Address for correspondence

Márcia Ito
 IBM Research Brazil
 Rua Tutóia, 1157 – 04007-900
 São Paulo/SP – Brazil
 e-mail: marciaito@br.ibm.com

Measurement of Respiration Rate and Depth Through Difference in Temperature Between Skin Surface and Nostril by Using Thermal Image

Hieyong Jeong^a, Yutaka Matsuura^b, and Yuko Ohno^{a, c}

^a Department of Robotics & Design for Innovative Healthcare, Osaka University, Suita, Osaka, Japan,

^b Graduate School of Dentistry, Osaka University, Suita, Osaka, Japan,

^c Division of Health Sciences, Osaka University, Suita, Osaka, Japan.

Abstract

The purpose of the present study was to propose a method to measure a respiration rate (RR) and depth at once through difference in temperature between the skin surface and nostril by using a thermal image. Although there have been a lot of devices for contact RR monitoring, it was considered that the subjects could be inconvenienced by having the sensing device in contact with their body. Our algorithm enabled us to make a breathing periodic function (BPF) under the non-contact and non-invasive condition through temperature differences near the nostril during the breath. As a result, it was proved that our proposed method was able to classify differences in breathing pattern between normal, deep, and shallow breath ($P < 0.001$). These results lead us to conclude that the RR and depth is simultaneously measured by the proposed algorithm of BPF without any contact or invasive procedure.

Keywords:

Thermography; Respiratory Insufficiency; Respiratory Rate.

Introduction

Respiratory failure is one of the most common reasons for admission to the intensive care unit (ICU) and a common comorbidity in patients admitted for acute care [1]. What's more, it's the leading cause of death from pneumonia and chronic obstructive pulmonary disease (COPD) in the United States. Patients with impending respiratory failure typically develop shortness of breath and mental-status changes, which may present as anxiety, tachypnea, and decreased peripheral oxygen saturation (SpO₂) despite increasing amounts of supplemental oxygen [2], [3]. It is possible to assess the patient's tissue oxygenation status regularly when the change of respiratory status is detected in the early stage. It follows that measuring the breath of the patient is important in the clinical environment.

The intensity of the care provided in the ICU requires many monitoring devices [4]. The ventilator, or respirator, is a breathing machine that helps patients breathe when they are too ill to breathe on their own. When signs and symptoms of respiratory failure are clearly strong, it is possible to use the ventilator at the appropriate time. However, when there are few signs and symptoms of respiratory failure, patients should be checked on a regular basis because symptoms can become suddenly or progressively worse. The current level of regular monitoring depends on the visual monitoring or stethoscope of caregivers, or the measurement of SpO₂ practically without any direct measurement of breath. In chronic respiratory failure, the

only consistent clinical indicator is protracted shortness of breath [5]. The ideal respiratory monitor will not only detect respiratory failure regardless of the cause, but also indicate the cause. Respiratory failure, which leads to severe medical problems, remains an important issue. However there seems to have been little to preclude our simple method of detecting the early stages of respiratory failure in the clinical environment.

Different respiratory monitoring methods measure different physiological variables [5]. According to our literature review, it is possible to divide existing non-invasive respiratory monitoring methods into three categories: detection of movement, volume and tissue composition; airflow sensing; and monitoring of blood gas concentrations.

Detection of movement, volume and tissue composition

Muscle activity causes variations in thorax volume and pressure that give rise to variations in the venous return of blood to the heart. Variations in thorax pressure cause variations in air volume within the lungs, which in turn result in variations in transthoracic impedance [6]. The methods included in this category are, besides transthoracic impedance monitoring [7], measurement of chest or abdominal circumference, electromyography, various motion detectors and photoplethysmography [8]. However, this method still struggles with measuring while patients have any movement.

Airflow sensing

It is possible to detect airflow because expiratory air is warmer, has higher humidity and contains more CO₂ than inspiratory air. Variations in these parameters can be used for indicating the respiratory rate. Even sound caused by the airflow can be used in a similar fashion. To be able to measure tidal volume, the airflow must be measured and integrated over time, or the volume itself must be measured. Both methods need a facemask or mouthpiece to collect the air. They are used in spirometers [9], but are not commonly used in monitoring devices [10]. The airflow sensing method has difficulties in measuring under the normal circumstances of daily life.

Monitoring of blood gas concentrations

Respiratory activity can be assessed indirectly by monitoring O₂ and CO₂ concentration in arterial blood using non-invasive methods. Oxygen saturation is available by pulse oximetry, end-tidal concentration of O₂ or CO₂ can be measured in expiratory air, and transcutaneous blood gas concentrations can be measured through the skin. When the end-tidal concentration of O₂ or CO₂ is measured, the relative change in concentration between expired and inhaled air for oxygen is less specific (roughly 21% for inhaled air and 16% for exhaled air) than that

for CO₂ (approximately 3% and nearly 0%, respectively) [3], [11], [12]. However, this method requires many contact points on the body of patient.

Purpose

From a present survey of the literature, it can be concluded that the need for effective respiratory monitoring is well known to the clinical engineering community. The number of papers in which new devices are demonstrated is large and, in fact, much larger than the number of publications devoted to the evaluation of such devices, indicating that the exploratory phase is still not finished [5]. Few solutions, if any, have been subjected to scrutinised evaluation on the basis of clinical usefulness. The reason for this may be that most monitoring devices are not yet adapted to clinical conditions. Monitoring respiratory activity in clinical practice introduces a number of problems that do not exist in a well-controlled laboratory-like situation. It must be possible to obtain the respiration status without disturbing or harming the patient. The monitoring devices should allow natural, unprovoked patient behaviour, such as turning to the side, moving head, arms and legs, and breathing through either nose or mouth [10]. Focusing on these reasons, the present study aims for caregivers to measure breathing patterns through the method related to airflow sensing under the non-invasive and non-contact conditions.

The main purpose of the present study was therefore to propose a method of measuring respiration rate (RR) and depth at once through differences in temperature between the skin surface and the nostril by using thermal imaging. Because our algorithm enabled us to make a breathing periodic function (BPF) under the non-contact condition during the breath, the peak-peak time and the distance between maximum and minimum peaks of BPF were helpful for measuring both the RR and depth at once.

Materials and Methods

Materials

Human subjects and experimental procedure

Three human subjects (Age: 29 ± 4 years old, Height: 171.4 ± 6.5 cm, Body mass index (BMI): 23.5 ± 3.2 kg/m²) were enrolled for the study at a local university. All of the subjects were informed of the nature of the experiments and consented for their data to be used in the study.

All thermal video recordings were carried out in parallel with a number of conventional contact-based respiration monitoring methods. This enabled a comparison to be carried out between the thermal imaging method and the conventional contact methods. The conventional contact respiration monitoring methods either measured the temperature variations between the inhaled and exhaled air using a nasal temperature probe, or measured SpO₂. The thermal video recordings were performed with the subject resting comfortably in a chair. The recordings did not cause any form of distress to the subject. The recording room temperature was about 24°C and humidity was about 30%.

Fig. 1 shows an experimental procedure in the present study. It is considered that there are three types of breathing: normal, deep, and intentional shallow breathing. We would like to confirm whether the proposed method can measure the RR and depth at once or not, while it does not matter if the type of breathing pattern is changed. A subject performed the normal breathing during 30 s, and then did two times deep breathing

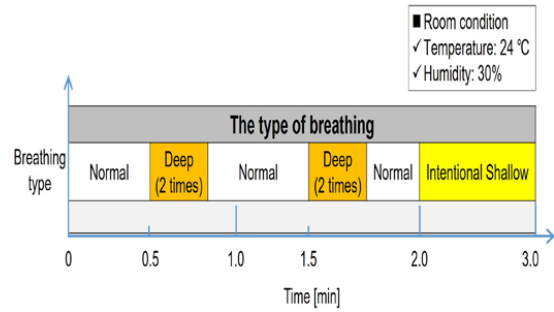


Figure 1 - Experimental procedure.

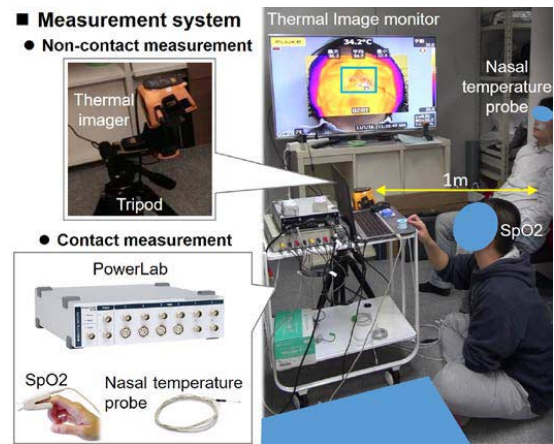


Figure 2 - Experimental system and environment.

within 1 min. Then the same performance was repeated one more time within an additional 1 min. Finally the subjects tried to perform the intentional shallow breathing during an additional 1 min, since for all subjects there were no problems with the experimental conditions. One subject performed three times under the same experimental conditions.

Experimental system

Fig. 2 shows an experimental system and environment. Apart from the 1.0 m distance, the thermal imager for the non-contact monitoring measures the ambient temperature of nostril and skin surface, and the nasal temperature probe for the contact monitoring also measures the ambient temperature around the nostril. The large monitor shows the respiration region of interest (ROI) which indicates the position of a rectangular boundary (blue rectangle) superimposed on a thermal image. A human pulse oximetry is used to measure blood oxygen saturation (SpO₂) in the finger. A computer saves and analyzes all measured data with the calculation and analysis software of Matlab.

The Fluke thermal imager (Ti450 60 Hz) with MultiSharp™ Focus [13] delivers focus near and far in one image. Focus is one of the most important aspects of thermography, and an out of focus image can yield data that can lead to misdiagnosis. The Ti450 brings 640 × 480 SuperResolution can show us even more detail with 4x the pixel data. An accuracy of temperature measurement was $\pm 2\%$ (at 25°C nominal, whichever is greater). The thermal imager could calculate each maximum, minimum and average value from the ROI automatically. The feature extracted from this ROI enabled the breathing signal to be produced. The thermal imager was fixed

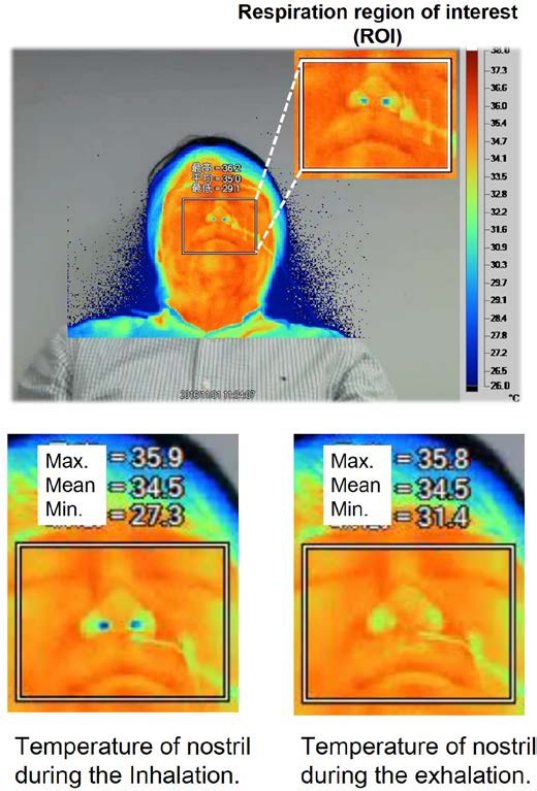


Figure 3 - The position of a rectangular boundary for the ROI superimposed on a thermal image (top) and differences in the temperature around nostril between inhalation and exhalation (bottom).

on a tripod in front of the subject at a distance of about one meter. Images were recorded at 9 frames per second (fps). The average respiration rate in an adult is about 15 cycles per minute (0.25 fps). Thus the image capture rate was sufficiently high. The duration of each recording was 3 min. This produced 1,620 thermal images per subject ($1 \times 9 \times 60 \times 3$). The recorded images were processed off-line using the Matlab image processing tool box in the present study. The operations involved for processing the images are explained in the next sections.

The Nasal Temperature Probe (MLT415/AL) contains a small biomedical chip thermistor [14]. The thermistor is designed to operate from 0°C to 50°C. The nasal temperature probe is suitable for connection to the ML309 Thermistor Pod. The accuracy of the sensor was $\pm 0.15^\circ\text{C}$ (if the temperature range was from 20°C to 35°C). The sampling frequency was set to 1 kHz.

Quick and easy to apply, these transducers (MLT321/MLT322 Human Pulse Oximetry) are ideal for spot checking and short-term continuous monitoring of SpO_2 [15]. The intensity of the transmitted light is measured and the ML320 Oximeter Pod produces a signal scaled to read in % SpO_2 . For this experiment, the finger clip was used. The accuracy of the sensor was $\pm 2\%$ of full scale for adults using the finger clip sensor. The sampling frequency was set to 1 kHz. The SpO_2 data was especially useful for confirming whether a subject performed the proper shallow breathing or not as shown in Fig. 1.

Methods

Fig. 3 (top) illustrates the position of a rectangular boundary for the ROI superimposed on a thermal image. To determine the respiration rate (RR) three phases are necessary: image acquisition, ROI tracking, and RR monitoring [16]. Image acquisition provides images from a thermal imager to ROI tracking and RR monitoring at a configuration rate. ROI tracking seeks the ROI in the current thermal image by providing a bounding box (BB). A BB consists of the coordinates (x, y) and the size (width $w = 7$ cm, height $h = 5$ cm) of the ROI. The size and position of BB can be changed by setting the thermal imager in advance. Although the BB coordinates are used to compute the position and orientation of the thermal imager, the BB is fixed to keep the ROI at the center of the image in the present study, as shown in Fig. 3 (top).

Finally, the RR can be calculated by the maximum and minimum temperature in the ROI. If the ROI is put to the m -by- n matrix of M_{BB} , both the maximum and minimum temperature are decided as follows:

$$\begin{aligned} T_{\max}(t) &= \max(M_{BB(m,n)}) [^\circ\text{C}] \\ T_{\min}(t) &= \min(M_{BB(m,n)}) [^\circ\text{C}] \end{aligned} \quad (1)$$

where the character of m represents rows of matrix M_{BB} , and the character of n represents columns. It is observed that mainly the maximum temperature in the ROI indicates the skin surface, and the minimum temperature in the ROI indicates differences in the temperature around the nostril during breathing as shown in Fig. 3 (bottom). Although the maximum temperature of skin surface in the ROI does not show much change during breathing, the minimum temperature of the nostril shows significant change between inhalation and exhalation. Thus the breathing periodic function (BPF) of the proposed algorithm is derived as follows:

$$BPF(t) = T_{\max}(t) - T_{\min}(t) [^\circ\text{C}] \quad (2)$$

The maximum peak value of BPF indicates breathing. The breath corresponds to the phases of inhalation and exhalation. It is considered that differences in temperature between inhalation and exhalation indicates breathing. The number of peaks indicates the number of breaths, and the time gap between one peak and the next peak in the BPF indicates the RR. Generally the RR is measured in number of breaths per minute (BPM), and the RR normally lies between 12 and 16 BPM at rest for an adult. Thus the RR in this paper is calculated as follows:

$$RR_{(i)} = \frac{60}{\text{time gap}} [\text{BPM}] \quad (3)$$

where the character of i represents the order of period of BPF.

Then, the distance gap between the maximum and minimum peaks indicates the depth of breathing. Thus the depth of breathing in this paper is calculated as follows:

$$\text{Depth}_{(in)} = \begin{cases} (\text{peak}_i^1 - \text{peak}_i^2), n = 1 \text{ if inhalation} \\ (\text{peak}_i^2 - \text{peak}_i^3), n = 2 \text{ if exhalation} \end{cases} \quad (4)$$

where the character of i represents the order of period of BPF. The meaning in $n = 1$ indicates the depth of breathing during the inhalation, and the meaning in $n = 2$ indicates the depth

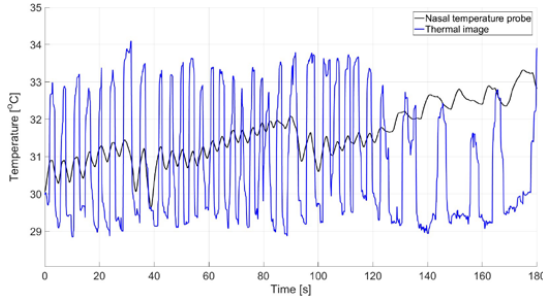


Figure 4 - Results of measured BPF by comparison with nasal temperature probe.

during the exhalation. It is considered that the distance gap of BPF means the relative depth of breathing. For example, the large difference in $Depth_{i1}$ or $Depth_{i2}$ by comparison with the normal breathing indicates deep breathing, and the small difference indicates shallow breathing.

Our proposed algorithm enables us to measure the RR and depth of breathing at once under the non-contact and non-invasive conditions by using the BPF of the thermal image as shown in equations (3) and (4). Furthermore, we would like to emphasize that it is possible to differentiate between inhalation and exhalation as shown in the equation (4).

Results

Results of measured BPF

It is necessary to understand the waveform of BPF which is used to measure the RR and depth of breathing. Figure 4 illustrates the results of measured BPF in comparison to measured data of the nasal temperature probe. The x axis indicates experimental time, and the y axis indicates temperature value. The black line represents the data of the nasal temperature probe, and the blue line represents those of BPF. The nasal temperature probe is regarded as the reference in the present study. In order to facilitate direct comparison between the two datasets, the initial value of BPF is added to the initial value of the nasal temperature probe. Because the initial value of BPF is about the lower value of 4.5°C before the addition of a baseline, it seems that it is difficult to compare differences between the two datasets directly. The increasing temperature value indicates the inhalation status, and the decreasing temperature indicates the exhalation status.

As a result, it was found that there were not significant differences in the period between the two different plots at all, although the type of breathing (normal, deep, and intentional shallow breathing) was changed according to the experimental procedure as shown in Fig. 1. Furthermore, although it seemed that there were big differences in maximum and minimum peaks between the two different data, it was no matter to extract featured points according to the different type of breathing.

The increasing slope of BPF after 2 min for the nasal temperature probe came from the abnormal weak inhalation and the influence of skin temperature on the measurement under the contact condition. However, because the proposed algorithm used the non-contact method, it was considered that the influence of the abnormal weak inhalation and skin temperature on the measurement was little.

Table 1 - Results of classification between the three different types of breathing.

No.	Index	The type of breathing		
		Normal	Deep	Intentional Shallow
■ Nasal temperature probe				
1	RR [BPM]	15.86 ± 2.17	10.51 ± 2.24	6.59 ± 2.47
2	Depth in inhalation [°C]	-0.39 ± 0.14	-1.11 ± 0.44	-0.43 ± 0.24
3	Depth in exhalation [°C]	0.44 ± 0.13	1.14 ± 0.47	0.66 ± 0.20
■ Thermal image				
1	RR [BPM]	15.99 ± 3.36	10.42 ± 2.47	6.61 ± 2.35
2	Depth in inhalation [°C]	3.88 ± 0.44	4.52 ± 0.41	3.02 ± 0.61
3	Depth in exhalation [°C]	-3.86 ± 0.43	-4.59 ± 0.44	-3.01 ± 0.52

Results of classification between the different types of breathing

It is necessary to prove whether it is possible to distinguish between the different types of breathing in the present study or not, through the proposed method. Table 1 shows the results of classification between the three different types of breathing. According to the experimental procedure with the three different types of breathing (normal, deep, and intentional shallow breathing), all of the indices are analyzed.

For the RR, values of 15.99 ± 3.36 BPM, 10.42 ± 2.47 BPM, and 6.61 ± 2.35 BPM for the thermal image under the condition of normal, deep, and shallow breathing were similar to those of 15.86 ± 2.17 BPM, 10.51 ± 2.24 BPM, and 6.59 ± 2.47 BPM for the nasal temperature probe under the same condition. This points to a high accuracy for the proposed algorithm for the measurement of RR. It was easy to distinguish the difference in the RR between the three different types of breathing ($P < 0.001$).

For the inhalation depth ($Depth_{i1}$), although values of $3.86 \pm 0.44^\circ\text{C}$, $4.52 \pm 0.41^\circ\text{C}$, and $3.02 \pm 0.61^\circ\text{C}$ for the thermal under the condition of normal, deep, and shallow were different from those of $-0.39 \pm 0.14^\circ\text{C}$, $-1.11 \pm 0.44^\circ\text{C}$, and $-0.43 \pm 0.24^\circ\text{C}$ for the probe under the same condition, it was easy to distinguish between the three different types of breathing (normal vs. deep: $P < 0.001$ and $P < 0.001$, normal vs. shallow: $P < 0.001$ and $P = 0.164$, deep vs. shallow: $P < 0.001$ and $P < 0.001$) except for the case (normal vs. shallow: $P = 0.164$ for the nasal temperature probe). As the possible explanation, it was considered that the increasing slope of BPF for the nasal temperature probe under the condition of shallow breathing as shown in Fig. 6 resulted in this problem. This may be why the weak inhalation under the condition of contact method resulted in the increasing slope of BPF due to the effect of skin surface temperature.

For the exhalation depth ($Depth_{i2}$), although values of $-3.86 \pm 0.43^\circ\text{C}$, $-4.59 \pm 0.44^\circ\text{C}$, and $-3.01 \pm 0.52^\circ\text{C}$ for the thermal under the condition of normal, deep, and shallow were different from those of $0.44 \pm 0.13^\circ\text{C}$, $1.14 \pm 0.47^\circ\text{C}$, and $0.66 \pm 0.20^\circ\text{C}$ for the probe under the same condition, it was easy to distinguish between the three different types of breathing (normal vs. deep: $P < 0.001$ and $P < 0.001$, normal vs. shallow: $P < 0.001$ and $P < 0.001$, deep vs. shallow: $P < 0.001$ and $P = 0.001$).

As a result, it was proved that our proposed method was able to classify differences in the RR and depth between normal, deep, and shallow breathing through the non-contact and non-invasive method by using the thermal image with a meaningful statistical significance under the natural conditions of daily life. In addition, because the BPF for measuring the RR and depth was based on differences in the temperature between the skin

surface and nostril, we would like to emphasize that it was considered that there was no difference among individuals. That was the reason why all subjects had differences in temperature between the skin and nostril during the breathing. These were strong points for the proposed method.

Conclusion

In this study, we demonstrated that the acquisition and analysis of the breathing signal are possible using thermal infrared imaging. The proposed algorithm, which measures the RR and depth of breathing through the BPF, is non-contact, non-invasive, and highly automated. This means that it can be safely employed to monitor patients in the clinical environment as well as at home for long periods of time.

The analysis of the experimental results indicates that the pattern of breathing is remarkably consistent for the same individual over time. Therefore, abnormalities in the individual patterns can be easily spotted. The feasibility of this new proposed algorithm is discussed for clinical studies.

Acknowledgements

This project was funded by the Hirose International Scholarship Foundation.

We wish to thank the industrial partners involved with the Konoike Institute of Technology, Konoike Transport Co., Ltd. (Room K708, 7th Floor, 3-1, Ofuka-cho, Kita-ku, Osaka 530-0011, Japan), who helped us find this avenue of research to begin with. We thank Mr. Tianyi Wang, Mr. Takafumi Ohno, Mr. Hideto Imai, and Prof. Michiko Kido, from Osaka University for their analysis of the measured data and their generous comments on the analysis results.

References

- [1] M. Fournier, Caring for patients in respiratory failure, *American Nurse Today* 9 (2014), 18-23.
- [2] K. Rehder and H. Michael Marsh, Respiratory Mechanics During Anesthesia and Mechanical Ventilation, *Handbook of Physiology, The Respiratory System, Mechanics of Breathing* (2011), 737-752.
- [3] A. Davies, A. G. H. Blakeley, C. Kidd, Human Physiology: The respiratory system, *Churchill Livingstone* (2001), 647-707.
- [4] G. S. Johnson, A Guide to ICU Equipment After Severe Brain Injury, *The Brain Injury Law Group* (Access date: 22 Nov 2016), <http://tbi-law.com/icu-equipment.html>.
- [5] M. Folke, L. Cernerud, M. Ekström, and B. Hök, Critical review of non-invasive respiratory monitoring in medical care, *Medical & Biological Engineering & Computing* 41 (2003), 377-383.
- [6] B. H. Brown, R. H. Smallwood, D. C. Barber, P. V. Lawford, and D. R. Hose, MEDICAL PHYSICS AND BIOMEDICAL ENGINEERING, *IOP Publishing Ltd.* (1999), Chapter 17-18.
- [7] K. P. Cohen, W. M. Ladd, D. M. Beams, W. S. Sheers, R. G. Radwin, W. J. Tompkins, and J. G. Webster, Comparison of impedance and inductance ventilation sensors on adults during breathing, motion, and simulated airway obstruction, *IEEE Trans Biomed Eng* 44 (1997), 555-566.
- [8] L. Nilsson, A. Johansson, and S. Kalman, Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique, *J Clin Monit Comput* 16 (2000), 309-315.
- [9] W. Perez and M. J. Tobin, Separation of factors responsible for change in breathing pattern induced by instrumentation, *J Appl Physiol* 59 (1985), 1515-1520.
- [10] M. Folke, F. Granstedt, B. Hök, and H. Scheer, Comparative provocation test of respiratory monitoring methods, *J Clin Monit* 17 (2002), 97-103.
- [11] M. W. Jopling, P. D. Manheimer, and D. E. Bebout, Effects of severe motion on three pulse oximeters designed for use in motion, *Int J Intens Care* (2000), 177-179.
- [12] T. E. Loughnan, J. Monagle, J. M. Copland, P. Ranjan, and M. F. Chen, A comparison of carbon dioxide monitoring and oxygenation between

facemask and divided nasal cannula, *Anaesth Intens Care* 28 (2000), 151-154.

- [13] FLUKE, (Access date: 22 Nov 2016), <http://en-us.fluke.com/products/infrared-cameras/fluke-ti450-infrared-camera.html>.
- [14] AD Instruments, Nasal Temperature Probes, (Access date: 22 Nov 2016), <https://www.adinstruments.com/products/nasal-temperature-probes>.
- [15] AD Instruments, SpO₂ Clips, (Access date: 22 Nov 2016), <https://www.adinstruments.com/products/spo2-clips>.
- [16] R. Chauvin, M. Hamel, S. Briere, F. Ferland, F. Grondin, D. Letourneau, M. Tousignant, and F. Michaud, Contact-Free Respiration Rate Monitoring Using a Pan-Tilt Thermal Camera for Stationary Bike Telerehabilitation Sessions, *IEEE Systems Journal* 10 (2016), 1046-1055.
- [17] C. L. Deepika, A. Kandaswamy, and G. Pradeepa, An efficient method for detection of inspiration phase of respiration in thermal imaging, *Journal of Scientific & Industrial Research* 75 (2016), 40-44.
- [18] R. Murthy and I. Pavlidis, Noncontact measurement of breathing function, *IEEE Eng Med Biol Mag* 25 (2005), 57-67.
- [19] A. K. Heiman, K. T. Orlikowsky, and L. Steffen, Non-contact respiratory monitoring based on real-time IR-thermography, *Proc World Congr Med Phys Biomed Eng* 25 (2009), 1306-1309.

Address for correspondence

Hieyong Jeong, PhD

Department of Robotics & Design for Innovative Healthcare,

Graduate School of Medicine, Osaka University

1-7 Yamadaoka, Suita, Osaka 565-0871, Japan

TEL: +81-6-6879-2614

E-mail: h.jeong@sahs.med.osaka-u.ac.jp

Visualizing Article Similarities via Sparsified Article Network and Map Projection for Systematic Reviews

Xiaonan Ji^{1,2}, Raghu Machiraju², Alan Ritter², Po-Yin Yen¹

¹Department of Biomedical Informatics, ²Department of Computer Science and Engineering,
The Ohio State University, Columbus, OH, USA

Abstract

Systematic Reviews (SRs) of biomedical literature summarize evidence from high-quality studies to inform clinical decisions, but are time and labor intensive due to the large number of article collections. Article similarities established from textual features have been shown to assist in the identification of relevant articles, thus facilitating the article screening process efficiently. In this study, we visualized article similarities to extend its utilization in practical settings for SR researchers, aiming to promote human comprehension of article distributions and hidden patterns. To prompt an effective visualization in an interpretable, intuitive, and scalable way, we implemented a graph-based network visualization with three network sparsification approaches and a distance-based map projection via dimensionality reduction. We evaluated and compared three network sparsification approaches and the visualization types (article network vs. article map). We demonstrated the effectiveness in revealing article distribution and exhibiting clustering patterns of relevant articles with practical meanings for SRs.

Keywords:

Information Storage and Retrieval; Data Display.

Introduction

Research findings in the biomedical literature are used to guide and improve clinical practice. Systematic reviews (SRs) summarize evidence drawn from high-quality and up-to-date studies to inform clinical decisions and are considered the preferred source of Evidence-based Practice (EBP) [1]. However, the number of studies published each week is over 12,000, including more than 300 randomized trials [2].

During the SR process, researchers conduct an exhaustive literature search and appraise the best evidence to answer a clinical question. With the overwhelming volume of studies (aka. articles), the article screening process has become the most burdensome aspect and often causes information overload [3]. Typically, an exhaustive literature search would yield hundreds or thousands of articles. SR researchers need to spend weeks or even months screening articles to identify relevant ones for inclusion. In general, only 2% to 30% of searched articles are included for evidence synthesis, which means researchers spend most of their efforts excluding irrelevant studies [4].

Existing studies have shown that automatic article classification with supervised machine learning (ML) is a valuable tool to facilitate the identification of relevant articles for SRs [4; 5]. However, such approaches have limited generalization to new SRs due to the dependency on prior supervised training data or manual annotations by domain experts. In our previous studies [6; 7], we demonstrated using established article similarities to assist in article screening for SRs in an unsupervised or semi-supervised manner. Article similarities were established in a

feature space derived from several article elements i.e., title and abstract. Our approach achieved competitive performance in reducing SR workloads, and is highly generalizable [6]. Article similarities were also shown to capture article distribution (the structure of an article collection) and the clustering of relevant articles based on their strong similarities [7].

To extend the utilization of article similarities, we proposed to delve into the visualization of article similarities, and compared the effectiveness of different visualization approaches in revealing article distribution and clustering patterns. Under the notion of information visualization [8], articles are represented as visual elements, and their similarities are encoded by connections or visual channels. As pictures can provide more information with less clutter in less space [9], transforming an article collection into graphical representations can enable human's insights into article distribution and clusters of similar articles. With the "visible article distribution", SR researchers can identify studies of interest more efficiently. While there are many visualization approaches [8], we considered that an effective visualization should display articles along with their similarities in a human interpretable, attribute intuitive, and spatial scalable manner [10]. In this study, we focused on two types of visualization: graph-based network visualization (article network) and distance-based (aka. geometry-based) map projection (article map). Other visualization types, such as adjacency matrix, hypergraph, and circular graph, were not included because of their limited structural analytics or spatial scalability.

For graph-based network visualization, we employed the node-link diagram with a force-directed layout to draw an article network, where articles are represented as nodes, and similarities are represented as weighted edges. Network visualization enables the exploration of graph topology and graph-based algorithms for advanced analytics [11]. However, an article network is almost a complete network due to the existence of non-zero similarities between most article pairs retrieved in an SR. Direct visualization of such a network is limited by human perception and important structural patterns are inaccessible, because of the extreme clutter presentation ("hairball") [12]. To provide an effective network visualization, we implemented three network sparsification methods to reduce the network size via edge sampling [13; 14], expecting to preserve edges bearing important conceptual or structural information. The three network sparsification methods include 1) the established article similarity (AS) that predominantly retains edges for strong article similarities, 2) the derived algebraic distance (AD), first proposed by Chen 2011 [15] and used in John 2016 [13], values edges within neighborhoods and tends to preserve a network's local structure, and 3)

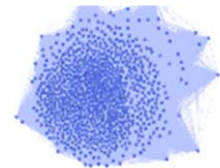


Figure 1. ADHD article

network visualization enables the exploration of graph topology and graph-based algorithms for advanced analytics [11]. However, an article network is almost a complete network due to the existence of non-zero similarities between most article pairs retrieved in an SR. Direct visualization of such a network is limited by human perception and important structural patterns are inaccessible, because of the extreme clutter presentation ("hairball") [12]. To provide an effective network visualization, we implemented three network sparsification methods to reduce the network size via edge sampling [13; 14], expecting to preserve edges bearing important conceptual or structural information. The three network sparsification methods include 1) the established article similarity (AS) that predominantly retains edges for strong article similarities, 2) the derived algebraic distance (AD), first proposed by Chen 2011 [15] and used in John 2016 [13], values edges within neighborhoods and tends to preserve a network's local structure, and 3)

the derived local degree (LD), first proposed by Lindner 2015 [14], values edges leading to hub nodes and tends to preserve the global structure. Both AD and LD have been shown to result in effective network sparsification [13; 14].

For distance-based map projection, we utilized the t-distributed Stochastic Neighbor Embedding (t-SNE) [16] to generate an article map with article features. The t-SNE technique converts the high-dimensional features into a matrix of pairwise similarities and visualizes it by projecting article points into a two-dimensional space, where article similarities are encoded as their spatial positions. t-SNE reveals local structures of the data and some important global structures such as clusters [16], though topological properties are not available in this map projection.

In summary, to effectively visualize article similarities, the purposes of the study are to 1) compare three network sparsification approaches, AS, AD, and LD; and 2) compare two types of visualization, graph-based network visualization (article network) and distance-based map projection (article map).

Methods

Dataset

We used publicly available data from 15 completed SRs [4; 17] produced by the Drug Effectiveness Review Project team (DERP). These SRs consist of article collections with coded decisions (inclusion or exclusion), which served as the gold standard to evaluate our performance. The size of these SRs ranges from 310 articles to 3465 articles (1249 articles on average), with the full-text level inclusion rate ranges from 0.55% to 27.04% (7.67% on average). We considered the full-text level included articles as relevant articles.

Article Features

Article similarities were established by article features. We generated (lexical) article features from several article elements, which were standardized by MEDLINE, including title (TI), abstract (AB), MeSH (MH), publication type (PT) and author (AU) [6]. We preprocessed the free-text of TI and AB by removing stop words, and stemming the remaining words with the classic Porter Stemmer. For MH, PT and AU, we used the exact strings as they were already standardly encoded. With the bag-of-words approach, we recorded the term frequency of each unigram word (TI and AB) and multi-word string (MH, PT and AU), and generated a feature vector for each article in the feature space. With article features, we calculated the article similarities to generate article networks. Alternatively, t-SNE calculated its own article similarities to generate article maps.

Article Network

An article network $G(V, E)$ consists of a set of nodes ($v \in V$) representing articles and a set of weighted edges ($e \in E$) representing article similarities between corresponding endnodes. We calculated article similarities (Euclidean distances) from article features using Cosine similarity. The resulting similarity ranges from 0 to 1 for each article element. We used an equally weighted sum of the five element similarities as the final article similarity for the edge weight, ranging from 0 to 5.

Network Sparsification

An article network is almost a complete network due to the existence of non-zero similarities between most article pairs in an SR. Visualizing such a network is meaningless as it is referred to as a “hairball”. On average, the article networks of 15 DERP SRs have approximately 1,200,000 edges. Figure 1 illustrates an article network using an SR, Attention Deficit Hyperactivity Disorder (ADHD), which has 851 articles.

To overcome the limitation of visualizing an almost complete network, we implemented three network sparsification methods, AS, AD, and LD, with an expectation to reduce the number of edges, but preserve edges with conceptual or structural importance. The sparsification process consists of three steps: (1) network pre-pruning, (2) edge scoring, and (3) edge sampling.

Step 1. Network pre-pruning. In our experiments, we found that directly calculating AD or LD edge scores using neighborhood information in an almost complete network resulted in indistinguishable edge scores as most nodes share similar neighbors. Therefore, we initiated a relaxed pre-pruning step to keep the top neighbors (edges) based on edge weights for each node. We tested a series of pre-pruning parameters from 50% to 5%, and found that keeping the top 10% edges led to a better balance of avoiding the mass of trivial edges and retaining important edges. At this step, the edge number has been reduced to 176,492 on average. We then used this coarsely pruned 10% network as the *baseline network*. The 10% baseline network, although it sounds satisfying, is still insufficient because the pre-pruning only cuts off very trivial edges. As most SRs contain more than 3,000 articles (some are more than 10,000) and result in a quadratic number of edges, network sparsification is needed to retain the most important edges. Thus in the following steps we further aggressively reduced the number of edges.

Step 2. Edge scoring. With the baseline network, we calculated edge scores to further capture the edge importance from different perspectives using AS, AD, and LD.

Article Similarity (AS) To aggressively retain edges corresponding to strong article similarities, we directly used the established article similarities as edge scores. In other words, the edge weights were used as edge scores.

Algebraic Distance (AD) AD was proposed to preserve strong connections in terms of local structures [13; 15]. It generalizes the idea of estimating the Jaccard coefficient for neighborhoods through lazy random walks to determine the strength of connections of the edges. Specifically, nodes with similar neighbors are considered strongly connected. With the AD approach, these nodes converge to similar values via information propagation within the neighborhood and lead to high AD edge scores between them. Thus, edges with high AD scores represent strong local connections. Algorithm 1 shows the computation of AD. We ran multiple rounds ($R=20$) to obtain synthesized results.

Algorithm 1 Computing algebraic distance (AD) [15]

Input: Parameter ω ($\omega = 0.5$), weighted adjacency Matrix (for weights w_{ij} and neighbors N_i), and randomly initialized vector $x^{(0)}$ with $|V|$ elements.

For $k = 1, 2$, do

$$\forall i \in V \ x_i^{(k)} \leftarrow \omega x_i^{(k-1)} + (1 - \omega) \frac{\sum_{j \in N_i} w_{ij} x_j^{(k-1)}}{\sum_{j \in N_i} w_{ij}}$$

$$\forall ij \in E \ s_{ij}^{(k)} = |x_i^{(k)} - x_j^{(k)}|$$

End for

Local Degree (LD) LD was proposed to emphasize “hub” nodes, which are nodes with relatively high degrees [14]. The hub nodes and the connections to the hubs are important to present a network’s global structure. Because LD used the unweighted degree, in our study, we extended it to the weighted degree for our weighted article networks. For each node, we scored an incident (associated) edge based on the weighted degree of the other endnode. The LD approach assigns high scores to edges that lead to the hub nodes, and preserves the network “hub backbone”.

Step 3. Edge sampling. After the edge scores were calculated, we sampled the edges based on the edge scores. For each

node $v \in V$, we included the top $\lceil \text{degree}(v)^e \rceil$ edges sorted by edge scores in descending order, where $\text{degree}(v)$ is the degree of node v , and e ($0 \leq e \leq 1$) controls the strength of sampling (filtering). We ensured that at least one incident edge was kept for each node. In this study, we used a sparsification parameter $e = 0.5$ to preserve at least $\lceil \text{degree}(v)^{0.5} \rceil$ edges for each node.

Force-directed Graph Drawing

We used a force-directed algorithm to draw sparsified article networks in an aesthetically pleasing way in a two-dimensional space. As a spatial layout, it places nodes and edges by simulating a physical system. When the system comes to a mechanical equilibrium state, the pairwise geometric distance between the drawn nodes matches the graph theoretic pairwise distance. Specifically, similar article nodes tend to aggregate together while dissimilar article nodes are drawn apart. We implemented the algorithm in Gephi with the built-in Force Atlas layout [18].

Article Map

In an article map, articles are represented as a set of points and their similarities are explicitly encoded as spatial positions of article points. We used t-SNE [16] to generate article maps by providing t-SNE the established article feature vectors. t-SNE establishes article similarities (Euclidean distances) in the high-dimensional feature space, and creates an article map by projecting article similarities down to a two-dimensional space.

t-SNE

t-SNE [16] is a technique for visualizing similarity data by embedding high dimensional data into a space of two or three dimensions. The resulting visualization is considered a map with data point distribution or a scatter plot. The t-SNE approach retains data structures by keeping similar data points close together while pushing dissimilar points far apart. It can also reveal important structures such as clustering.

As a non-linear algorithm for dimensionality reduction, t-SNE establishes high-dimensional Euclidean distances (similarities) between data points and converts them into a probability distribution. In the low-dimensional space, map points are placed as counterparts with a similar probability distribution. Gradient descent is used to minimize the divergence between the two distributions with respect to map points' spatial positions. It is worth mentioning that the gradient of the cost function can be interpreted as physical forces between map points just like the force-directed graph drawing for networks. t-SNE has been shown to create higher-quality visualizations than linear methods (i.e. PCA and MDS) and other nonlinear methods (i.e. SNE and Isomap) [16]. We implemented t-SNE in MATLAB.

Evaluation

Network Properties Evaluation

We evaluated article networks based on the network properties in a network structure, including graph diameter, clustering coefficient, communities, and modularity. *Graph diameter* is the length of the shortest path between the most distanced nodes. A smaller diameter indicates a stronger concentration of a graph. *Clustering coefficient* measures the degree to which nodes tend to cluster together. Nodes with a higher clustering coefficient have higher transitivity in the neighborhood. *Communities* are subsets of nodes that are internally densely connected but externally sparsely connected. *Modularity* is designed to measure the strength of division of a network into communities. A high modularity corresponds to a better community structure. We used the Louvain method [19] for community detection which is proven to provide high-quality results. A graph's community structure and modularity also reflect its local structure with respect to intra-community connections. In addition, a graph's

global structure can be reflected by the diameter, averaged clustering coefficient, and the number of communities. However, these topological properties were unavailable in article maps.

Clustering Patterns Evaluation

We evaluated the clustering patterns on both article networks and article maps. For article networks, we utilized communities detected by the Louvain method. For article maps, we applied k-means clustering to identify clusters based on the 2-dimensional map computed by t-SNE. For convenience, we used *set* to refer to *community* and *cluster*. We evaluated the clustering patterns of relevant articles that have been identified in the completed DERP SR reports (external criteria). Because of the highly-imbalanced dataset with only 0.55%–27.04% (7.67% on average) of relevant articles, we identified the *set* that contains at least 10% of relevant articles as *dominant set*. We examined the coverage (*recall*) and proportion (*precision*) of relevant articles in all dominant sets, and calculated the balanced F-measure (*F1 score*). This was inspired by the classic measure of clustering quality that evaluates how well the clustering matches the gold standard classes, and interprets the clustering as a series of decisions. Other measures such as the purity, normalized mutual information, and Rand index, are not suitable for the highly-imbalanced dataset.

Results

Network Properties

We reported the network properties of the baseline network and the sparsified networks with the averaged results of the 15 DERP reports (Table 1). Because edge sampling was implemented based on individual nodes, the number of edges after the sampling was not the same when using different sparsification methods. The clustering coefficient was also the average of all article nodes for each SR.

Table 1- Preservation of network properties

	Base-line	AS	AD	LD
Edge Number	176,492	12,706	13,323	15,722
Diameter	3	6	5	4
Clustering Coefficient	0.4038	0.2581	0.1169	0.3806
Modularity	0.4103	0.6231	0.4336	0.3860
Community Number	5	10	7	5

As shown in Table 1, the number of edges was significantly reduced from 176,492 to approximately 13,000 after sparsification. All sparsification methods resulted in increased diameters and decreased clustering coefficients because of the removal of most edges. However, they also showed differences. AS brought the highest modularity with a larger number of communities, but altered the baseline graph diameter and clustering coefficient to a greater extent. Similarly, AD altered the baseline diameter and clustering coefficient, but provided a slight gain in modularity. LD retained a similar graph diameter, clustering coefficient, modularity, and community number compared to the baseline. In summary, AS resulted in a better community structure; AD tended to retain the baseline local structure (with slightly higher modularity); LD performed the best in preserving the baseline global structure.

Clustering Patterns

We examined the clustering patterns of relevant articles in article networks and article maps. The optimized number of communities in an article network was determined by the *Louvain method* (default resolution setting). For the *k-means* clustering in an article map, we applied the *knee (elbow)*

method to identify a proper value range for the number of clusters, k . We found the resulting range approximately aligned to the number of communities detected by the Louvain method in AS networks. Thus for each SR, we had k equal to the number of communities in the corresponding AS network.

In Table 2, we reported the total number of sets (communities or clusters), the number of relevant sets that contain at least one relevant article, and the number of dominant sets that contain at least 10% of relevant articles. We also reported the overall size of all dominant sets by calculating the ratio of articles contained by the dominant sets. We calculated the corresponding recall, precision, and F1 score regarding the relevant articles in all dominant sets. Again, all the results were averaged from the 15 DERP SR reports.

Table 2- Article distribution and clustering of relevant articles

	Article Networks				Article Map
	Base-line	AS	AD	LD	
Total Set #	5	10	7	5	10
Relevant Set #	4	6	5	4	8
Dominant Set #	3	3	3	3	3
Dominant Size	63.10%	35.13%	56.35%	65.59%	31.36%
Recall	94.87%	83.16%	90.14%	92.99%	76.45%
Precision	10.84%	17.53%	12.40%	10.08%	17.36%
F1 Score	0.1823	0.2623	0.1983	0.1713	0.2618

As shown in Table 2, the AS network and article map generated by t-SNE had the largest number of sets (10) and relevant sets (6 and 8), but the number of dominant sets was only 3. Their dominant sets covered 83.16% and 76.45% of relevant articles with a size of 35.13% and 31.36% of entire articles. Both of them had lower recalls but the highest precisions and F1 scores (0.2623 and 0.2618). They achieved a good quality of clustering relevant articles by decomposing articles into finely separated sets. The LD network behaved similarly to the baseline network, with the relevant articles spreading into coarsely divided sets. With the highest recalls, 3 out of 5 sets acted as dominant sets and covered over 90% of relevant articles. However, their precisions and F1 scores (0.1713 and 0.1823) were lower than others because the conservative discrimination. The AD network brought moderate performance in recall, precision, and F1 (0.1983). Specifically, 90.14% of relevant articles were covered by dominant sets, with a size of 56.35%.

In summary, network sparsification led to a more recognizable network structure by distributing articles into distinct sets. According to the F1 score, the quality of clustering relevant articles was improved with AS and AD. In addition, the AS network and article map had the highest precisions and F1 scores, but the AS network had a better recall than the article map. Overall, the sparsified article networks and the article map provided a more interpretable distribution and clustering patterns.

Illustration of Visualization

We illustrated the above mentioned results with visualization using an SR report, ADHD, as an example. The ADHD report has a total of 851 articles: 20 were included at the full-text level (relevant articles), 64 were only included at the title/abstract level, and 767 were excluded at the title/abstract level.

Figure 2 shows article networks before and after sparsification. The baseline network had clutter presents that limited human perception without explicit structures. All sparsified networks provided more interpretable structures and revealed the clustering of relevant articles (green nodes). Specifically, AS provided the most manifest community structure with meticulous separations, where relevant articles were highly

concentrated. AD retained local connections and led to densely connected neighborhoods. LD preserved the hub backbone structure by concentrating nodes towards hubs and forming bridges among hubs. Figure 3 shows an article map generated by t-SNE. We observed the clustering patterns of relevant articles (green points).

In Figure 4, we illustrated communities in the AS network and clusters in the article map. Article nodes (points) were colored by communities (clusters). Dominant sets of relevant articles were marked by green rectangles, which further demonstrated the effective clustering of relevant articles.

Discussion

Sparsification Schemes: AD and LD were applied to sparser networks (i.e. social networks and citation networks) in early works [13; 14]. To our knowledge, we were the first to apply AD and LD to article network sparsification. Due to the densely-connected nature of article networks, a relaxed pre-pruning step was used. In this study, we found that AD retained the local network structure, LD preserved the global network structure (also supported by other works [13; 14]), and AS performed the best in revealing the community structure. In addition, considering the clustering of relevant articles, AD and LD had lower precisions, but higher recalls, resulting from their integration of the network structure; while AS had a lower recall, but a higher precision because it aggressively concentrated relevant articles. Another encouraging finding was that by keeping only 1-3% of edges (13,000 edges on average) from the original networks (1,200,000 edges on average), we can reveal meaningful network structures and important clustering patterns.

Article Network vs. Article Map: Article networks sparsified by AS and articles maps generated by t-SNE achieved similar results in revealing article distribution and clustering patterns. Both of them aggregated the majority of relevant articles into finely separated set(s). While the AS network had a slightly lower precision than the article map, it had a higher recall which is important for SRs. For article networks, a sparsification process is needed to eliminate clutter presents; but we were able to explore graph topology and apply graph-based algorithms for advanced visual analytics, such as community detection and graph traversal. Article maps are created by t-SNE or other dimensionality reduction algorithms that can handle the crowded article feature space; but topological analysis is not available. The same article feature generation step was applied to the network visualization and the map projection process. Future investigation could include a user study to gather feedback regarding the selection of visualization approaches for SRs.

Significance for SRs and Future Work: Effective visualization of article similarities has practical and significant implications for SRs. By revealing the structure and distribution of articles, SR researchers can conduct SRs in a more effective and efficient manner. Consider a scenario of screening hundreds of articles in the ADHD report, researchers can rapidly identify relevant articles by screening only a small percent of articles once the dominant sets are located. In our future work, we will further investigate the identification of the dominant set. We also plan to provide visual feedback on the selection of article features, and integrate advanced semantic features to improve the similarity calculation.

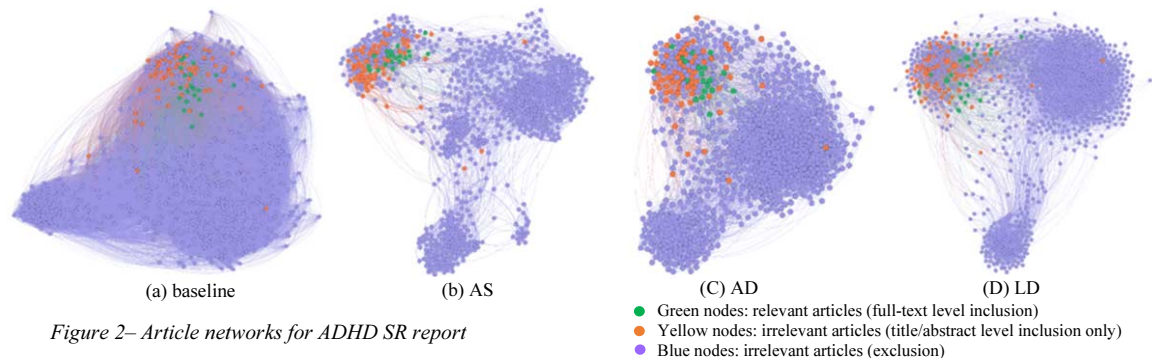


Figure 2—Article networks for ADHD SR report

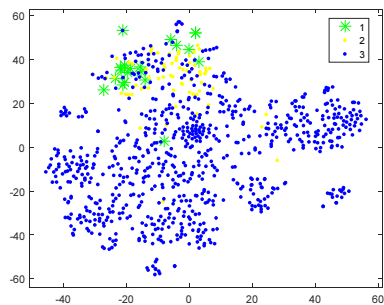


Figure 3—Article map for ADHD SR report

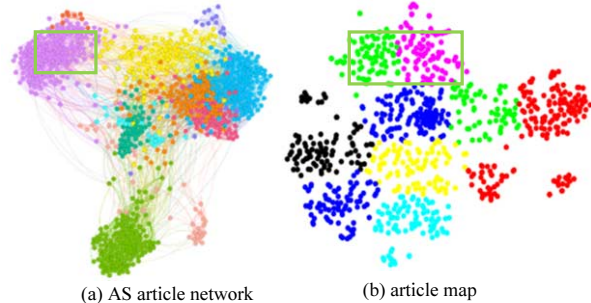


Figure 4—Dominant sets of relevant articles (ADHD)

Conclusions

We visualized article similarities with sparsified article networks and article maps. We demonstrated the effectiveness in revealing meaningful articles structures, and exhibiting clustering of relevant articles in an intuitive and human interpretable manner. Effective visualization of article similarities has practical meanings to facilitate article screening for SR researchers.

Acknowledgements

This work was supported by the Agency for Healthcare Research and Quality (AHRQ), R03HS025047-01.

References

- [1] Oxford Centre for Evidence-based Medicine - Levels of Evidence. <http://www.webcitation.org/6AZnaE7iG>
- [2] P.P. Glasziou, Information overload: what's behind it, what's beyond it, *Medical Journal of Australia* **189** (2008).
- [3] S. Majid, S. Foo, B. Luyt, X. Zhang, Y.-L. Theng, Y.-K. Chang, and I.A. Mokhtar, Adopting evidence-based practice in clinical decision making: nurses' perceptions, knowledge, and barriers, *Journal of the Medical Library Association* (2011).
- [4] A.M. Cohen, W.R. Hersh, K. Peterson, and P.-Y. Yen, Reducing workload in systematic review preparation using automated citation classification, *Journal of the American Medical Informatics Association* **13** (2006), 206-219.
- [5] S. Matwin, A. Kouznetsov, D. Inkpen, O. Frunza, and P. O'Brien, A new algorithm for reducing the workload of experts in performing systematic reviews, *Journal of the American Medical Informatics Association* **17** (2010), 446-453.
- [6] X. Ji and P.-Y. Yen, Using MEDLINE Elemental Similarity to Assist Article Screening for Systematic Reviews, *JMIR* (2015).
- [7] X. Ji, R. Machiraju, A. Ritter, and P.-Y. Yen, Examining the Distribution, Modularity, and Community Structure in Article Networks for Systematic Reviews, in: *AMIA*, 2015.
- [8] R. Lengler and M.J. Eppler, Towards a periodic table of visualization methods for management, in: *LASTED Proceedings of the Conference on*

Graphics and Visualization in Engineering (GVE 2007), Clearwater, Florida, USA, 2007.

- [9] A. Kerren, J. Stasko, J.-D. Fekete, and C. North, *Information Visualization: Human-Centered Issues and Perspectives*, Springer, 2008.
- [10] C. Ware, *Information visualization: perception for design*, Elsevier, 2012.
- [11] H. Ding, C. Wang, K. Huang, and R. Machiraju, GRAPHIE: graph based histology image explorer, *BMC Bioinformatics* **16** (2015), 1.
- [12] P. Hu and W.C. Lau, A Survey and Taxonomy of Graph Sampling, *arXiv preprint arXiv:1308.5865* (2013).
- [13] E. John and I. Saffro, Single- and Multi-level Network Sparsification by Algebraic Distance, *arXiv:1601.05527 [cs.SI]* (2016).
- [14] G. Lindner, C.L. Staudt, M. Hamann, H. Meyerhenke, and D. Wagner, Structure-Preserving Sparsification of Social Networks, *arXiv:1505.00564v1 [cs.SI]* (2015).
- [15] J. Chen and I. Saffro, Algebraic distance on graphs, *SIAM Journal on Scientific Computing* **33** (2011), 3468-3490.
- [16] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, *Journal of Machine Learning Research* **9** (2008), 85.
- [17] Drug Effectiveness Review Project (DERP) Systematic Drug Class Review Gold Standard Data. skynet.ohsu.edu/~cohenaa/systematic-drug-class-review-data
- [18] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software, *PLoS One* **9** (2014), e98679.
- [19] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment* **2008** (2008), P10008.

Address for correspondence

Xiaonan Ji: ji.62@osu.edu
1800 Canon Drive, Columbus, OH 43210, USA

A Hybrid Method for ICD-10 Auto-Coding of Chinese Diagnoses

Zheng Jia^a, Weifeng Qin^a, Huilong Duan^a, Xudong Lv^a, Haomin Li^{b,c}

^a College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

^b The Children's Hospital, Zhejiang University, Hangzhou, China

^c The Institute of Translational Medicine, Zhejiang University, Hangzhou, China

Abstract

The Chinese Version of Classification and Codes of Diseases (CCD) is an expanded version of ICD-10. Hospitals are required to assign CCD codes to discharge diagnoses in China. To handle the contradiction between a shortage of skilled CCD coders and increasing coding efficiency, a CCD auto-coding method is urgently needed. In this study a hybrid auto-coding method was proposed based on the lexical characteristics obtained through the analysis of a corpus of 1537 diagnoses with normative CCD code. It combines the rule-based approach, the Chinese characters-based distributed semantic similarity and the dictionary-based approach. The rule-based approach was proved to be efficient and precise at the cost of time and manpower. The semantic similarity approach shows poor performance. The old-fashioned dictionary-based approach ends in leading significance. The final accuracy of this hybrid approach is 96.9% in the test.

Keywords:

International Classification of Diseases; Clinical Coding; China

Introduction

The International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) [1] was defined as a system of categories to which morbid entities are assigned according to established criteria. The ICD was used to translate diagnoses of diseases and other health issues from linguistic words into alphanumeric codes, which permits easy storage and retrieval, and systematic recording analysis of mortality and morbidity data. The classification consists of 22 chapters, which were divided into homogeneous blocks of categories, which are further subdivided to at most 10 subcategories each.

The Chinese Version of Classification and Codes of Diseases (GB/T 14396-2001) [2], short for CCD, is an expanded version of ICD-10 and is widely used in China. It has a strict and organized classification hierarchy expanded from four-char ICD-10 (9505 subcategory) to six-char codes (23106 expansion) that allows one-to-one mapping from diseases to codes to support the management, inquiry, search and statistical analysis of data.

In 2011, the National Health and Family Planning Commission of the People's Republic of China promulgated the notice on using CCD codes for encoding discharge diagnoses in the summary page of medical records after January 1, 2012 [3].

Accurate CCD diagnosis coding is critical to patient care, billing purposes, and research endeavors [4]. Correct, standard and complete hospitalization diagnosis by doctors and careful implementation of coding principle and careful reading about

medical record content are the guarantee of ensuring correct coding [5]. The patient's course of disease is defined as a procedure from admission to discharge with a narrative discharge summary including the primary and secondary diagnoses by manual input. The paper trail is defined as a creation of medical record from the recording of the admitting diagnosis to the assignment of the ICD codes after discharge [6][6].

Errors such as "upcoding" and misspecification may be introduced during the coding by the lacking of medical knowledge and coder's experience, attention, and persistence, the ability to interpret the diagnoses and the inconsistency between CCD phraseology and clinical phraseology, such as the synonyms and abbreviations used to describe the same condition. [6].

In most of large hospitals, several professional coders are hired by the medical record department to supplement the CCD codes in the summary page of hospital records [7]. Nevertheless, the contradiction between a shortage of skilled coders and an excess of hospital patients makes it a tough task to encode both the primary and the secondary discharge diagnoses with ICD-10. In a first-class hospital in China only two coders are hired to assign CCD codes for about 2000 inpatients everyday.

Physicians in the same department usually follow the same naming convention so that there is little difference in the record names of the same disease in its summary page of hospital records. Similar grammar mistakes and medical language process makes it feasible to analyze the underlying rules of encoding diagnoses into CCD codes and develop a CCD auto-coding tool for physicians.

To reduce the workload of human coders, Medori and Fairon [8] extracted necessary information from French discharge summaries and combined the symbolic approach and statistical approach, since a large corpus of clinical notes was available. Subotin and Davis [9] developed a system for predicting ICD-10-PCS codes from the English clinical narrative using partial hierarchical classification and confidence calculation and estimation. In China, Ning and Yu [10] proposed an algorithm to implement ICD-10 coding automatically for clinical diagnoses in Chinese and calculated the semantic similarity between terms by the definition of distributed semantic similarity. Considering the linguistic features of Chinese, the results indicate that term vectors constructed from words have a higher precision than that from Chinese characters.

In this paper, we analyze and summarize the characteristics of discharge diagnoses written by doctors in the Nephrology Department, Dayi Hospital. A rule-based data pre-processing method and a computer-assisted CCD coding method are proposed.

Methods

In this study, we collected a corpus, analyzed the lexical characteristics of Chinese diagnoses and proposed three approaches to auto-code CCD.

Corpus Collection

The corpus of patient Electronic Medical Record (EMR) summary pages were collected from a Clinical Data Repository (CDR) [11] system, which was implemented in the 2000-bedded hospital in China. The summary page was in XML format and each summary page recorded one primary discharge diagnosis of the patient and its associated CCD code assigned by human coders. A corpus of 1537 diagnoses with CCD codes considered as the gold standard was randomly selected and parsed from data repository as the typical language samples to analyze the lexical characteristics.

Lexical Characteristics Analysis

Due to individual language habit, doctors may use diverse names to define the primary diagnoses and describe the status of disease [6]. ICTCLAS [12], which is a Chinese lexical analyzer, is used to tokenize the long disease names into minimal semantic units (Table 1).

Table 1 – Result of Tokenization

Original Name	Tokenization
慢性肾衰竭急性加重 chronic kidney failure exacerbate acutely	慢性肾衰竭/n 急性 /b 加重/n
轻度系膜增生性IgA肾病 mild membranoproliferative IgA nephropathy	轻度/d 系膜增生性 IgA肾病/n
无症状性血尿 asymptomatic haematuria	无/v 症状/n 性/ng 血尿/n
不典型膜性肾病 atypical membranous nephropathy	不/d 典型/a 膜性 肾病/n

Tags: n – Noun, b – distinguishing words, v – verb, d – adverb, a – adjective, ng – Noun morpheme.

Table 3 – Rules for CCD Auto-coding

Pattern	Original Name	Processed Name
prefix term	<u>右</u> 肾结核 <u>right</u> kidney tuberculosis	肾结核 kidney tuberculosis
term suffix	腰椎管狭窄 <u>症</u> lumbar spinal stenosis <u>disease</u>	腰椎管狭窄 lumbar spinal stenosis
term1 and/with term2	肾病综合征 <u>合并</u> 急性肾损伤 nephrotic syndrome <u>and</u> acute injury of kidney	肾病综合征 nephrotic syndrome 急性肾损伤 acute injury of kidney
term1 punctuation1 term2 punctuation2	肾炎综合征, <u>膜</u> 性肾病 nephrotic syndrome_ <u>membranous</u> nephropathy	肾炎综合征 nephrotic syndrome 膜性肾病 membranous nephropathy
character in similar form	肾病综合 <u>症</u> nephrotic syndrome	肾病综合 <u>征</u> nephrotic syndrome
homophone	急进 <u>性</u> 肾炎综合征 rapidly progressive nephritic syndrome	急进 <u>性</u> 肾炎综合征 rapidly progressive nephritic syndrome
synonymous symbol	Roman numeral ⇔ Arabic numeral	
synonymous words	甲状腺 <u>功能</u> 减退症 hypothyroidism	甲状腺 <u>机</u> 能减退症 hypothyroidism
remove non-Chinese character	腰5骶1椎间盘突出 lumbar 5 sacrum 1 disc herniation	腰骶椎间盘突出 lumbar sacrum disc herniation

The analysis results include the tokenization result and observational conclusions (Table 2). The hand-input phrases usually consist of three components: multiple attributes, one core noun and punctuations. The punctuations consist of redundancy such as bracket and digit symbols. Several common wrongly written characters of discharge diagnoses are discriminated and the reason is analysed.

Table 2 – Semantic Class of Chinese Diagnoses

Semantic class	Examples
Attribute	<u>弥漫性</u> <u>膜性</u> 肾小球肾炎 <u>diffuse membranous</u> glomerulonephritis
Noun	弥漫性膜性 <u>肾小球肾炎</u> <u>diffuse membranous glomerulonephritis</u>
Punctuation	肾炎综合征, <u>膜性</u> 肾病 nephrotic syndrome_ <u>membranous</u> nephropathy 肾炎综合征 (<u>膜性</u> 肾病) nephrotic syndrome (<u>membranous</u> nephropathy)
Errors	homophone, Chinese character in similar form, synonymous symbol, synonymous words, et al.

Computer-assisted Coding Method

We proposed three approaches for automated CCD code assignment: the rule-based approach, the Chinese characters-based distributed semantic similarity approach and dictionary-based approach.

Rule-based auto-coding

The rule-based approach contains nine irregular naming patterns established from practice and experience (Table 3).

Superfluous prefix or suffix are omitted. Typical prefix words include anatomic parts (eg. “下肢” Chinese for “leg”), nouns of locality (eg. “左” Chinese for “left”) and scope (eg. “多发” Chinese for “multiple”), words end with particular Chinese characters (eg. 性 property).

Conjunctions (eg. “合并” Chinese for “accompanied”) and meaningful punctuations(eg. “逗号” Chinese for “comma”) are sometimes used to connect two individual disease names. The CCD dictionary may exclude the combination but include the individual.

Each permutation of the rules applied to one original diagnosis name will generate a candidate. The candidates that can be matched to any disease name in CCD dictionary or custom dictionary (much more on this later) completely will be selected. The standard disease name will also be processed if it contains particular characters or symbols. The rule-based auto-coding is invalid if it's not an exact match. The candidate with a perfect match and maximum character count will be recommended as the paired CCD code.

Chinese characters-based distributed semantic similarity

The Chinese characters-based distributed semantic similarity [10] is a method which can implement ICD-10 coding automatically for clinical diagnoses in Chinese and has a high precision in the test set. It is measured by the cosine similarity between vectors converted from Chinese characters.

$\vec{p} = (p_1, p_2, \dots, p_M)$ and $\vec{q} = (q_1, q_2, \dots, q_N)$ are defined as the standard CCD name (the count of Chinese characters is m) and the hand-input name (the count is n) respectively.

$\vec{p}' = (p_{i_1}, p_{i_2}, \dots, p_{i_K})$, $\vec{q}' = (q_{j_1}, q_{j_2}, \dots, q_{j_K})$ are the vectors extracted from \vec{p} and \vec{q} where $p_{i_k} = q_{j_k}$ ($k \in [1, K]$), $\vec{p}' = \vec{q}'$, $p_i \in \{p_1, p_2, \dots, p_m\}$, $q_j \in \{q_1, q_2, \dots, q_n\}$. The similarity score is

$$\text{sim}(p, q) = \frac{K}{M} \cdot \frac{K}{N} \cdot |\vec{p}' - \vec{q}'|$$

where

$$\vec{p}' = (i_1, i_2, \dots, i_K), i_k \in [1, M]$$

$$\vec{q}' = (j_1, j_2, \dots, j_K), j_k \in [1, N]$$

and M is the count of Chinese characters of the standard CCD name, N is the count of Chinese characters of the hand-input name, K is the count of common characters.

For example, there are one original disease name 巨幼细胞贫血(megaloblastic anaemia) and three candidates of CCD name 巨幼红细胞性贫血(megaloblastic anaemia), 巨幼细胞遗传性贫血(megaloblastic hereditary anaemia), 营养性巨幼细胞性贫血(nutritional megaloblastic anaemia). We have

$$\vec{q} = (\text{巨}, \text{幼}, \text{细}, \text{胞}, \text{贫}, \text{血}), N = 6$$

$$\vec{p}_1 = (\text{巨}, \text{幼}, \text{红}, \text{细}, \text{胞}, \text{性}, \text{贫}, \text{血}), M_1 = 8$$

$$\vec{p}_2 = (\text{巨}, \text{幼}, \text{细}, \text{胞}, \text{遗}, \text{传}, \text{性}, \text{贫}, \text{血}), M_2 = 9$$

$$\vec{p}_3 = (\text{营}, \text{养}, \text{性}, \text{巨}, \text{幼}, \text{细}, \text{胞}, \text{性}, \text{贫}, \text{血}), M_3 = 10$$

The extracted vectors are $\vec{p}_1' = \vec{p}_2' = \vec{p}_3' = \vec{q}' = (\text{巨}, \text{幼}, \text{细}, \text{胞}, \text{贫}, \text{血})$, $K = 6$, where the sequence vectors and similarity scores are

$$\vec{q}' = (1, 2, 3, 4, 5, 6)$$

$$\vec{p}_1' = (1, 2, 4, 5, 7, 8), \text{sim}(p_1, q) \approx 2.37$$

$$\vec{p}_2' = (1, 2, 3, 4, 8, 9), \text{sim}(p_2, q) \approx 2.83$$

$$\vec{p}_3' = (4, 5, 6, 7, 9, 10), \text{sim}(p_3, q) \approx 4.95$$

The non-Chinese characters are removed before similarity calculation. The candidate code with the lowest similarity score will be assigned to the disease name. In the above example, the CCD code of p_1 should be assigned to q .

Dictionary-based approach

The dictionary-based approach is to construct a custom dictionary which extended the CCD standard disease name with disease names used by doctors and its corresponding CCD

codes given by human coders. The scope of the dictionary tends to expand over time. The term absented from CCD dictionary will be assigned to its newly assigned CCD code when it appears again.

The three approaches are integrated to follow the priority order: custom dictionary > rule-based auto-coding > Chinese characters-based distributed semantic similarity. If the original diagnosis fails to match any CCD of the standard CCD dictionary or the custom CCD, it is then processed by the rules. This procedure may generate several candidates. If no candidate is obtained, the auto-coding fails. Otherwise the candidates are compared with CCD of the standard CCD dictionary or the custom CCD. If no candidate matches successfully, the CCD code with a highest score of Chinese characters-based distributed semantic similarity is selected as the final result. In a situation where there are multiple matches in standard or extended CCD dictionary, the first candidate in the default sort is selected as the final match. The flow diagram of the hybrid method is shown in Figure 1.

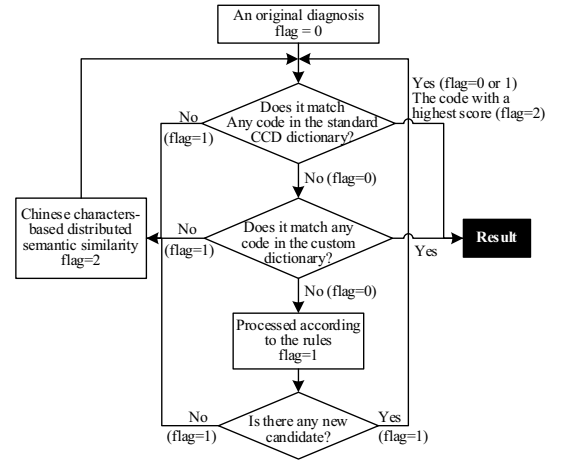


Figure 1 – The Flow Diagram of the Hybrid Method

Results

The corpus of 1537 diagnoses were auto-coded according to the hybrid approach. Then we verified the validity of recommended CCD codes by the gold standard codes assigned by human coders. In Table 4 the count of right results of each pattern was calculated programmatically. Each pattern contributes to a bit of the final result and the percentage number indicated the contribution.

About ten percent of raw diagnoses written by doctors could be mapped to one CCD code without any medical language processing. It indicated that the language habit and common parlance in the clinical context varied among terms in CCD dictionary.

Rule-based auto-coding was an effective method with high accuracy but low recall. We applied regex functions for specified patterns; that is, candidates were searched among both the standard disease names and the original disease names according to the rules. It supplemented 42.4% diagnoses with correct CCD codes. We developed and reviewed every rule carefully at the cost of time and manpower. A further investigation was carried out and it came to the conclusion that the auto-coding rules were largely consistent across all departments.

Table 4 – Auto-coding Result

Approach	Pattern	Count	Percent
Direct Match		161	10.5%
Processed Standard Name	term1 punctuation1		
	term2 punctuation2	285	
	prefix term suffix	12	
	prefix term	114	
	term suffix	27	
	term1 and/with term2	4	
Processed Original Name	term1 punctuation1		
	term2 punctuation2	123	42.4%
	character in similar, form, homophone, synonymous words	18	
	synonymous symbol	57	
	remove non-Chinese character	12	
Chinese characters-based distributed semantic similarity		18	1.2%
Dictionary-based		659	42.9%
Mismatch		47	3.1%
Count		1537	100%

Chinese characters-based distributed semantic similarity method was a method dedicated to Chinese characters with mediocre accuracy, poor efficiency and small influence. Only 1.2% discharge diagnoses were under valuable influence.

The dictionary-based approach improved the matching rate from 52.9% to 97%. It proved to be an invalid approach initially and would play a leading role as it scaled. It may start out as unreliable, but it can be trained and developed into a practical handbook coders can rely on. At present, most of the hospitals seldom established this kind of dictionary and even had no information system assisting the human coders [13].

For mismatched cases, some errors are analyzed. One case is that the original name is 双下肢动脉硬化伴多发斑块形成 (the hardening of the arteries of both lower limbs with multiple plaque buildup) and the standard name is 下肢动脉粥样硬化 (atherosclerosis of arteries of lower limbs). Another case is that the original name is 系统性小血管炎肾损害 (kidney damage due to (not mentioned in Chinese) systemic small-vessel vasculitis) and the standard names are 系统性血管炎 (systemic small-vessel vasculitis) and 系统性结缔组织疾患引起的肾小球疾患 (glomerular disorders in systemic connective tissue disorders, ICD-10: N08.5*). It infers that mismatches may be caused by the usage of totally different syntax.

Discussion

The combination of the three approaches reached 96.9% precision for the auto-coding of 6-char CCD in the Nephrology Department. Ning et al. [14] used a hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation, reaching the precision of 4-char coding is about 92%. It indicates that our method improved the fineness without diminution of accuracy.

The rule-based method and the dictionary-based approach cost running times within 1 second, while the Chinese characters-based distributed semantic similarity is long-running process. The average running time of the method proposed by Ning [14] is about 1 second. For the little sense made by the Chinese characters-based distributed semantic similarity method, it is suggested to be abandoned.

The rule-based method in this paper is essentially an extended dictionary matching approach enriched with rule pre-processing. It is not a traditional rule based approach for other tasks. A hospital-wide, highly efficient, precise auto-coding system coordinating the rule-based method and the dictionary-based approach is the next challenging task in our project.

The rule-based method and dictionary-based approach take merely a fraction of operation time. However, the Chinese characters-based distributed semantic similarity method requires more internal storage and longer execution time, because the dictionary is large and the similarity calculation and comparison between each code is time-consuming.

Conclusion

In this paper we introduced the usage condition of ICD-10 in China and the expanded version of ICD-10 known as CCD. The existing approaches assisting human coders to assign CCD codes to diagnoses context were reviewed. Three methods were proposed. Rule-based auto-coding method is of high effectiveness and precision. Chinese characters-based distributed semantic similarity method performed poorly. The dictionary-based approach proved to be gradually crucial and reliable.

However, there are still limitations. The hybrid approach is of low portability and interoperability, because rules for auto-coding and the custom dictionary differ from hospital to hospital. Rule-making and dictionary construction for each single hospital is of high cost and low efficiency. In future work, we will conduct the in-depth analysis of auto-coding and develop a hospital-wide system to assist the work of human coders.

Acknowledgements

This research was financially supported by the national key research project 2016YFC0901905.

References

- [1] W.H. Organization, *International statistical classification of diseases and health related problems (The) ICD-10*, World Health Organization, 2004.
- [2] General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, *Classification and Codes of Diseases*, in: Standards Press of China, 2002.
- [3] Notice of the General Office of the Ministry of Health on the Issues concerning Classification and Codes of Diseases (Revision), 2011.
- [4] C.P. Chen, S. Braunstein, M. Mourad, I.-C.J. Hsu, D. Haas-Kogan, M. Roach, and S.E. Fogh, Quality improvement of International Classification of Diseases, 9th revision, diagnosis coding in radiation oncology: Single-institution prospective study at University of California, San Francisco, *Practical radiation oncology* **5** (2015), e45-e51.
- [5] C. Li, X. Chen, An Analysis on Reasons of Wrong Coding for 5000 Cases of Medical Records, *Chinese Medical Records* **12** (2011), 23-24.
- [6] K.J. Omalley, K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, and C.M. Ashton, Measuring Diagnoses: ICD Code Accuracy, *Health Services Research* **40** (2005), 1620-1639.
- [7] Y. Wang, N. Guo, X. Bai, Discussion on Coder Cultivation in County-Level Hospitals, *Chinese Medical Record*, **12**(2015):81-82.
- [8] J. Medori and C. Faron, Machine learning and features selection for semi-automatic ICD-9-CM encoding, in: *Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents*, Association for Computational Linguistics, 2010, 84-89.
- [9] M. Subotin and R. Davis, A system for predicting ICD-10-PCS codes from electronic health records, *ACL 2014* (2014), 59.
- [10] X. Ning, M. Yu, Algorithmic Research on Automatic Coding of Clinical Diagnoses Based on Semantic Similarity Calculation, *Journal of Medical Informatics*, **37** (2016), 52-56.
- [11] L. Min, L. Wang, X. Lu, and H. Duan, Case Study: Applying OpenEHR Archetypes to a Clinical Data Repository in a Chinese Hospital, in: *MEDINFO 2015: EHealth-enabled Health: Proceedings of the 15th World Congress on Health and Biomedical Informatics*, IOS Press, 2015, p. 207.

- [12] P. Zhang, K. Yu, Y. Xiong, and Q. Liu, HHMM-based Chinese lexical analyzer ICTCLAS, in: *Proceedings of the second SIGHAN workshop on Chinese language processing-Volume 17*, Association for Computational Linguistics, 2003, 184-187.
- [13] L. Yang, M. Yu, An Application of a Computer Aided ICD-10 Coding System, *Chinese Medical Record*, (2015), 28-32.
- [14] W. Ning, M. Yu, and R. Zhang, A hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation, *BMC Medical Informatics and Decision Making* **16.1** (2016), 30

Address for correspondence

Zheng Jia
College of Biomedical Engineering and Instrument Science,
Zhejiang University, Hangzhou, China.
Email: jiaz@vico-lab.com
Tel: +8615858296829

The Impact of “Possible Patients” on Phenotyping Algorithms: Electronic Phenotype Algorithms Can Only Be Reproduced by Sharing Detailed Annotation Criteria

Rina Kagawa, MD^a, Yoshimasa Kawazoe, MD, PhD^b, Emiko Shinohara, PhD^b,
Takeshi Imai, PhD^c, Kazuhiko Ohe, MD, PhD^{a,b}

^a Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Japan

^b Department of Healthcare Information Management, The University of Tokyo Hospital, Japan

^c Center for Disease Biology and Integrative Medicine, The University of Tokyo, Japan

Abstract

Phenotyping is an automated technique for identifying patients diagnosed with a particular disease based on electronic health records (EHRs). To evaluate phenotyping algorithms, which should be reproducible, the annotation of EHRs as a gold standard is critical. However, we have found that the different types of EHRs cannot be definitively annotated into CASEs or CONTROLS. The influence of such “possible patients” on phenotyping algorithms is unknown. To assess these issues, for four chronic diseases, we annotated EHRs by using information not directly referring to the diseases and developed two types of phenotyping algorithms for each disease. We confirmed that each disease included different types of possible patients. The performance of phenotyping algorithms differed depending on whether possible patients were considered as CASEs, and this was independent of the type of algorithms. Our results indicate that researchers must share annotation criteria for classifying the possible patients to reproduce phenotyping algorithms.

Keywords:

Clinical Phenotyping, Data Annotation, Electronic Health Records

Introduction

Background and problems

To improve healthcare quality and clinical research, it is critical to identify patients diagnosed with a particular disease. As structured data on diagnoses in electronic health records (EHRs) are limited in terms of accuracy and completeness [1, 2], the demand for automated techniques for identifying patients diagnosed with a particular disease based on EHRs, so-called phenotyping, has been increasing [3, 4]. In previous studies, we applied published algorithms to EHR datasets of Japanese patients [5, 6], which required annotated EHRs as a gold standard to evaluate the algorithms. Among the several annotation techniques, we chose manual annotation because it is more accurate than others, and over 75% of previous studies employed it [4]. Through the annotation process, we found that it is often difficult to annotate EHRs as definite CASE or CONTROL patients as a gold standard [6].

Annotation criteria are critical for phenotyping algorithms because they directly affect the calculation of the algorithms' performance. If researchers do not share how to annotate such “possible patients,” the published performance of a phenotyping algorithm would differ among research teams because the CASEs would differ, even if each dataset had identical characteristics (Figure 1).

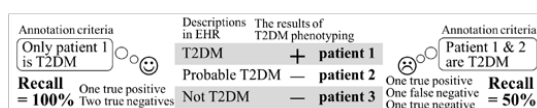


Figure 1—The performance of phenotyping algorithm (recall) for classifying T2DM cannot be reproduced across institutions due to the different annotation criteria for classifying the “possible patient (patient 2).”

Why is the annotation of possible patients indefinite?

In EHR annotation for type 2 diabetes mellitus (T2DM), we identified three types of EHR data in terms of possible patients [6]. Each of these suggests a different likelihood of T2DM, as shown in the following examples:

Example 1: Antigliutamic acid decarboxylase (GAD) antibody ≥ 1.6 U/mL and antiislet antigen 2 (IA2) antibody ≥ 0.4 U/mL

Example 2: “...T2DM is likely...”

Example 3: “...He met the diagnostic criteria for DM... Type 1 DM is unlikely because... Secondary DM is denied...”

In example 1, the explicit information about the disease name, such as the direct noting of “type 1 DM,” is not provided, but the information implies a low possibility of T2DM; that is, the high value of the anti-GAD antibody or anti-IA2 antibody suggests type 1 DM. Example 2 includes an explicit but indefinite description (“likely”), and the annotation result may differ among annotators. Example 3 provides no direct information about the type of DM, but the contextual information can increase the conviction that the patient has T2DM; that is, it denies other types of DM. These ambiguous descriptions are sufficient for medical experts who have medical knowledge, and they sometimes dare to describe EHRs ambiguously to accurately record the facts when they cannot diagnose patients with certainty [7]. However, these descriptions are not necessarily sufficient for researchers who sometimes expect the definite truth. This is one of the essential limitations of retrospective EHR-based studies across institutions or countries. Moreover, the fact that each patient's EHRs contain many such ambiguous descriptions makes reproducible annotation difficult.

As surveyed, one study separated definite and possible CASEs to identify rheumatoid arthritis and grouped possible CASEs with CONTROLS [8]. For other diseases, such as multiple sclerosis, the possible patients were identified for multiclass classifications [9–11]. However, they have not shown multiple types of possible patients, and the published annotation criteria were disease dependent. For other diseases, even the existence of possible patients as a gold standard has not been examined. No studies have shown whether the performance of a

phenotyping algorithm will be influenced by which types of possible patients are classified as CASEs. We aim to quantitatively clarify these issues and mitigate the ambiguities to facilitate reproducible phenotyping algorithms [12].

Research objective, novelty, and related work

We analyze EHR data to examine the impacts of the annotation criteria for classifying possible patients on (1) the proportion and characteristics of CASE patients and (2) the performance of phenotyping algorithms. To accomplish this, we create a nonbinary annotation method based on the conviction of a target disease. Our targets are four chronic diseases. This study’s originality is to handle directly the uncertainties in records of diseases and their influences on the algorithms’ performance using real world data. One study simulated the loss of power of phenotyping algorithms due to bias in the EHR data [13], but did not mention the information bias related to the annotation criteria. No phenotyping studies have addressed the uncertainties in disease records [7].

Methods

Target diseases, eligible subjects, and EHR data

Our targets are four diseases from the Unified Medical Language System (UMLS) Metathesaurus—common diseases: (i) T2DM and (ii) essential hypertension (HT); and rare diseases: (iii) primary biliary cirrhosis (PBC; the full disease name is changed since approximately 2015 [14], but this study used data until 2014) and (iv) autoimmune hemolytic anemia (AIHA). We selected chronic diseases, which have associated clinical guidelines in Japan [15–18], to focus only on the presence of diseases. A disease must be diagnosed through a combination of several tests. If a disease can be diagnosed using one test, it does not require a phenotyping algorithm. This study involved 650 patients (mean age 52.6; 57.2% female) randomly selected out of 104,522 patients who made at least two visits to the University of Tokyo Hospital between 1/1/2009 and 12/31/2014 and at least one visit in 2012. We used the EHR data over six years (2009–2014).

Detailed annotation (DA) method

In this study, annotators checked EHR data retrospectively and determined a “CASE” based on the degree of conviction that the patient had a target disease, which was recorded in EHRs by clinicians who examined the patient. The examples of EHRs in the Introduction section suggest that possible patients should be annotated into several types and that annotation should be independent of the data structure. Labelling the information itself would be useful for reproducible annotation of any EHRs because the annotation results would differ among research teams using different information. We divided the information in the EHRs into two axes, namely explicit information and context (Figure 2); moreover, along each of the two axes, we classified the elements affecting multilevel degree of conviction of a target disease. We have called this the detailed annotation (DA) method (Table 1).

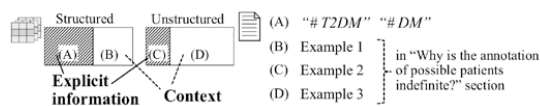


Figure 2—Annotation axes: explicit information and context.

Explicit information includes definite (Table 1(a), (d)) and possible ((b), (e)) descriptions of disease names, and meeting the diagnostic criteria ((c), (e)), which means that annotators can retrospectively determine that the patient met the

diagnostic criteria. The description “T2DM is likely” is an explicit, possible description of the target disease (T2DM) and belongs to explicit information (b). Contexts include the EHR data implying a target disease (α); upper diseases (β); and the absence ($(\alpha), (\beta)$) and presence (ϵ) of differential diagnoses and sibling diseases; and possible (δ) and definite (ϵ) descriptions of diseases which are treated by the same medication used for the target disease. For T2DM, the description “Type 1 DM is unlikely” belongs to context (α) because T1DM is a sibling disease of T2DM; this description implies T2DM. The pair of the structured data of the high values of the anti-GAD antibody and anti-IA2 antibody belongs to context (ϵ) because they suggest T1DM. For essential HT, the narrative and definite descriptions that the patient is medicated with an antihypertensive drug not for HT but for angina belongs to (ϵ); this description implies weak conviction that the patient has essential HT.

The combination of each element of each of the two axes leads to the seven categories of disease conviction, from a definite CASE (category (1)) to a definite CONTROL (category (7)). As an intermediate concept, category (4) indicates the upper disease. A patient with explicit information (b) and context (α) is classified into DA category (2). Patients in the categories (2)–(6) are possible patients.

Experimental setups

Two clinicians (authors) each annotated the EHRs of all 650 patients for each of the four diseases based on the DA categories. The annotators discussed and made final decisions for mismatches. We used Fisher’s exact test and Fisher’s pairwise exact test (Bonferroni correction) to compare the proportion of possible patients among the diseases. To analyze the patients’ characteristics according to the DA categories, we performed statistical analyses of the null hypothesis that the averages of the maximum value of each lab test of each patient or the proportions of each element used in the guideline-based phenotyping algorithms (described later) would be equal among the categories. For continuous data, the variances of each category were found to be unequal by a Bartlett test; the nonparametric Kruskal–Wallis test and the Mann–Whitney U test (Bonferroni correction) for pairwise comparisons were performed. For categorical data, we used Fisher’s exact tests and Fisher’s pairwise exact test (Bonferroni correction). The threshold for significance was $p < 0.05$. We used R-3.1.3, the coin package 1.1-2, and the fmsb package 0.5.2.

To assess the impact of the annotation criteria on the performance of phenotyping algorithms, authors developed rule-based phenotyping algorithms for each disease (Table 2) using data that are referred in each guideline [15–18] and are stored according to HL7 2.5 (ISO 27931:2009). We call these guideline-based (GB) algorithms. For each algorithm, we changed the annotation criteria by altering the threshold of the DA category in which patients are considered as CASEs from the category (1) to (6), and calculated the values of the evaluation metrics at each threshold. The DA category (7) was excluded from the threshold because it was an unrealistic assumption for all patients (the categories (1)–(7)) to be CASEs. The evaluation metrics are as follows: recall is True Positive (TP)/(TP+False Negative (FN)), precision was TP/(TP+False Positive (FP)), specificity was True Negative (TN)/(TN+FP), and negative predictive value (NPV) was TN/(TN+FN). To examine whether the impact of the annotation criteria depended on a particular phenotyping algorithm, we performed the same experiment for the other phenotyping algorithms, which used only one billing code for each disease. We call these naïve algorithms. More detailed explanations of the DA method and phenotyping algorithms are presented on GitHub.¹

Table 1– The DA method. Degrees of disease conviction are classified into categories (1)–(7) indicating certainty, probability, possibility, upper disease, the possibility of other diseases, the probability of other diseases, and definite other diseases respectively.

Context		(α) Increasing the conviction of the target disease	(β) Probable upper disease AND no sibling diseases or differential diagnoses	(γ) No other context	(δ) Possible diseases which are treated by the same medications used for the target disease	(ε) Decreasing the conviction of the target disease	(ζ) Definite denial of the target disease
Explicit information							
Diagnostic criteria	Description						
Target disease name	(a) Definite	(1)	(1)	(1)	(1)	(3)	Contradiction
	(b) Possible	(2)	(3)	(4)	(5)	(5)	(7)
	(c) Meet	(3)	(4)	(4)	(5)	(6)	(7)
Upper disease name	(d) Definite	(2)	(4) or (6) ²	(4) or (6) ²	(5)	(6) or (7) ²	(7)
	(e) Meet	(3)	(4) or (6) ²	(4) or (6) ²	(5) or (6) ²	(6) or (7) ²	(7)
Both target and upper	(f) None of the others	(5)	(6)	(7)	(7)	(7)	(7)
	(g) Denial	Contradiction	(7)	(7)	(7)	(7)	(7)

Table 2– GB phenotyping algorithms developed in this study.

(i) T2DM: Patients with (A) AND ((B) OR (C)), modified [19]	
(A)	Excluding other types of DM
(B)	Antidiabetic medication
(C)	T2DM billing codes AND abnormal lab test more than two times.
(ii) Essential HT: Patients with (A) AND (B)	
(A)	Excluding secondary HT
(B)	HT billing codes except (A) OR medication of ARB/ACE inhibitor
(iii) PBC: Patients with (A) AND (B)	
(A)	Antimitochondrial antibody (AMA) is positive
(B)	(γGTP ≥ 68 IU/L AND ALP ≥ 359 IU/L) OR PBC billing codes
(iv) AIHA: Patients with ((A) AND (B)) OR (C)	
(A)	More than four abnormal lab tests, which mean hemolytic anemia
(B)	Direct Coombs test is positive AND (AIHA billing codes OR no billing codes for other diseases that cause anemia)
(C)	Disease names of AIHA in EHR, applying the technique in [20]

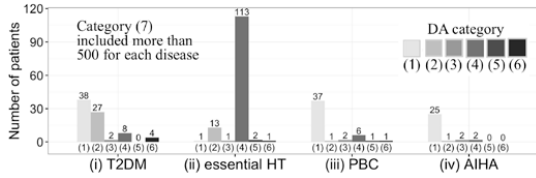


Figure 3– The numbers and distributions of patients classified into each DA category differ among the diseases.

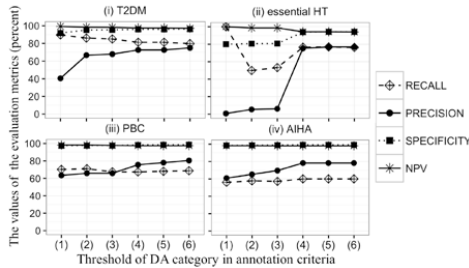


Figure 4– Changes in the performance of GB algorithms depending on the threshold of CASEs differ among the diseases.

Results

Figure 3 shows the annotation results based on the DA categories (weighted κ statistics = 1.00 for each disease). The mismatches between the annotations of two patients was caused by an oversight of the EHRs and those of seven patients were caused by misunderstanding of the contexts. The category (7) included more than 500 patients for each of the four diseases and is excluded from Figure 3. All four diseases included possible patients classified into DA categories (2)–(6). The distributions of patients differed among the diseases. T2DM had a broad peak at the categories (1) and (2), and essential HT had a peak at upper disease (the category (4)). PBC included patients in all DA categories, while AIHA did

not include patients in the categories (5)–(6). The proportions of possible patients to all 650 patients were not equal among the diseases ($p < 2.2e-16$, two sided). Post-hoc test showed significant differences between all disease pairs ($p < 0.001$) except the pair of PBC and AIHA ($p = 0.14$).

If the characteristics of patients in the DA categories (1)–(6) were exactly equal, the degrees of disease conviction recorded by the clinicians in EHRs would be completely random, and the classification of possible patients would not be necessary. We performed statistical analyses to assess this. Because the category (7) indicates definite CONTROLS, it will naturally have different characteristics from the other categories and was excluded from the statistical analysis. The averages of the maximum HbA_{1c} of each patient differed significantly among the DA categories (1)–(6) (Table 3(a)). Post-hoc tests showed no significant difference for each category pair (Table 3(b)). The other lab tests' values did not differ significantly. The proportions of tests for glucose ($p = 0.048$), tests for anti-GAD antibody ($p = 0.021$), antidiabetic medication ($p = 0.0009$), insulin ($p = 0.040$), T2DM billing codes ($p = 0.009$), and T1DM billing codes ($p = 0.0002$) differed significantly (two sided) among the categories (1)–(6). A post-hoc test showed no significant differences. Although not significant, the other categorical variables tended to exhibit higher proportions in the category (1) than that in the categories (2)–(6) collectively, i.e., PBC billing code and no billing codes for malignant neoplasms that cause anemia (regarding AIHA, Table 3(c)).¹²

Figure 4 shows that the performance of each GB phenotyping algorithm changed depending on the threshold of the DA category in which patients were considered as CASEs. The values of the evaluation metrics at the threshold DA category (2) (hereinafter called th(2)) indicated the values when the definite CASEs (category (1)) and probable CASEs (category (2)) were considered as CASEs. For essential HT, the recall decreased by 50% as the threshold moved from th(1) to th(2) (Figure 4(ii)). It is because that the GB algorithm(ii)(A) excluded 5 out of 13 patients in DA category (2) who had secondary HT billing codes, and that the GB algorithm(ii)(B) also excluded two patients in the category (2) (Table 2). The precision and the specificity increased by 69.2% and 13.1% from th(3) to th(4), respectively; these changes were not found in other diseases. They depended on the 90-patient increase in TP patients and concomitant decrease in FP patients, which were due to the peak at category (4) (Figure 3(ii)). In contrast, as the threshold moved from th(1) to th(6), the decrease in the NPV of PBC (0.66%) or AIHA (0.16%), and the increase in the specificity of AIHA (0.63%) were within 1% (Figure 4(iii), (iv)). PBC and

¹ <https://github.com/rinabouk/medinfo2017>

² T2DM and essential HT are smaller in number; they account for more than 50% of upper diseases. PBC and AIHA are larger in number; they account for less than 50%.

AIHA included significantly fewer patients in categories (2)–(6). From th(1) to th(6), the changes in the numbers of TP, FP, FN, and TN patients and the changes in performance were smaller than for common diseases. The changes for T2DM were intermediate; the recall and NPV decreased by 9.73% and 2.12%, respectively, and the precision and the specificity increased by 39.3% and 4.49%, respectively (Figure 4 (i)). Figure 5 shows that the changes in performance of naïve algorithms, which exhibited the same patterns as in Figure 4 except for the decrease in recall from th(1) to th(2) for essential HT.

Table 3(a)– The averages of the maximum HbA_{1c} value of each patient were not equal in the DA categories (1)–(6).

	(1)	(2)	(3)	(4)	(6)	p-value
Maximum value of	7.78	7.15	8.20	6.76	6.20	0.007
HbA_{1c} of each patient	(1.21)	(0.85)	(1.40)	(0.16)	(0.22)	
average (SD)	(n=35)	(n=22)	(n=2)	(n=5)	(n=3)	

Table 3(b)– No significant differences between any two category pairs for the maximum HbA_{1c} .

	p-value	Effect size		p-value	Effect size
(1) (2)	0.28	0.20 (small)	(2) (4)	1.00	0.29 (small)
(1) (3)	1.00	0.02 (no)	(2) (6)	0.17	0.56 (large)
(1) (4)	0.33	0.42 (medium)	(3) (4)	1.00	0.82 (large)
(1) (6)	0.08	0.57 (large)	(3) (6)	0.75	0.97 (large)
(2) (3)	1.00	0.14 (small)	(4) (6)	0.25	0.70 (large)

Table 3(c)– Proportions of billing codes for malignant neoplasms that cause anemia in category (1) (11/14) tended to be lower than in categories (2)–(6) (3/1).

	(1)	(2)	(3)	(4)	(5)	(6)	p-value
Patients with billing codes	11	1	0	2	0	0	0.21
Patients with no billing codes	14	0	1	0	0	0	

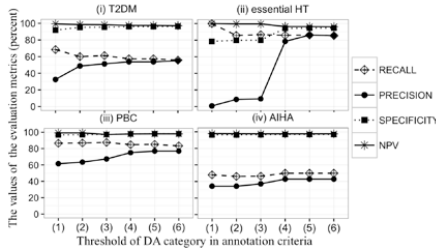


Figure 5– Pattern of changes in the performance of naïve phenotyping algorithms. These are the same as in Figure 4.

Discussion

Our experiments showed that for each of the four chronic diseases, it was necessary to determine how to annotate CASEs by dividing possible patients from definite CASEs and classifying each type of possible patients, to avoid research teams’ inclusion of different characteristics in their CASEs. Consequently, the changes in performance of phenotyping algorithms following the alteration in annotation criteria differed among the diseases; this was considered to be independent of the type of algorithms, that is, GB or naïve algorithms. We suggest that these results support the importance of sharing annotation criteria in detail to reproduce algorithms.

The characteristics of possible patients

The different distributions of patients among the diseases (Figure 3) are clinically plausible. Clinicians can diagnose T2DM willingly by a combination of simple lab tests or symptoms [15]; many clinicians were assumed to diagnose

T2DM with strong conviction (category (1)) or describe information that inferred T2DM (category (2)). Essential HT is a diagnosis by the exclusion of secondary HT [16]; most clinicians only describe “HT” when referring to essential HT. This was assumed to be the reason for the peak at category (4). PBC and AIHA are rare diseases; then, specialists diagnose most patients with strong conviction (category (1)) [17, 18]. Further, the upper disease of AIHA (acquired hemolytic anemia) is rare, while that of PBC (fibrosis and cirrhosis of liver) cannot be sometimes diagnosed with certainty even by specialists because it is relatively common; then, PBC had a relatively higher proportion of possible patients compared to AIHA, and AIHA did not include the patients in categories (5) and (6).

The detailed annotation criteria according to the degrees of disease conviction recorded by the clinicians reflect patient characteristics and clinicians’ rational assessments (Tables 3(a)–(c)). We considered that clinicians could neither diagnose patients with low HbA_{1c} certainly nor describe the definite disease name for such patients, and the average of the maximum HbA_{1c} of each patient in DA categories (4) or (6) for T2DM was relatively low. Similarly, it is suggested that the clinicians’ choices of lab tests, medications, or billing codes are affected by the clinicians’ conviction of the diseases. It seems one reason no elements tended to be unequal among the categories for essential HT was that they included antihypertensive medications for other diseases and the corresponding billing codes, independent of the conviction of essential HT.

Annotation criteria influence the reproducibility of phenotyping algorithms

Many studies have applied phenotyping algorithms to multiple institutions and identical performance has not been achieved [19, 21]. Our findings suggest that even if the research teams used the same annotation criteria, the ambiguities of the criteria and the corresponding different annotation results could lead to the different performance. This is a critical limitation when other teams attempt to reproduce published performance for different datasets. Without detailed annotation criteria, other teams could not judge whether the different performance arises from differences in annotation criteria, the different characteristics of the study population, or the differences in available data or tools. Our results can guide researchers on this limitation; the robust metrics in terms of changing annotation criteria, such as the NPV of rare diseases, would be preferable for reproducing algorithms without shared annotation criteria. In addition, failure to share annotation criteria could lead to erroneous interpretations of published performance. For example, the precisions of the GB algorithm for T2DM were 39.3% different between th(1) and th(6) (Figure 4(i)).

We confirmed that these issues do not depend on the particular phenotyping algorithm; although only for essential HT, the decrease in recall by 50% from th(1) to th(2) was dependent on the GB algorithm. One simulation study showed that different gold standards led to the different sensitivity or specificity in diagnostic studies [22]. Our results were consistent with this, and presented the first assessments of the different gold standards in studies for phenotyping algorithms using actual data. If all patients diagnosed with a certain disease had definite descriptions of the disease in their EHRs, no patients would be in the categories (2)–(6). Each evaluation metric would retain the same value independent of our annotation criteria, and there would be four flat lines in Figure 4 or 5. The diseases examined in this study were not such cases. Higher recalls of essential HT and PBC, and precision of essential HT could be achieved using naïve algorithms, while the scope of this study was the change in performance. Which algorithm is better is to be determined based on each research objective or available data.

Because the annotation criteria are to be determined according to the research purposes, they will naturally differ among studies. Thus, sharing annotation criteria in detail is critical in reproducing EHR-based studies.

Limitations and future work

To assess our results' generalizability, we must evaluate them with patients with other diseases at other hospitals in several countries. Lower weighted κ statistics would be obtained from this study. Nevertheless, when applying the same phenotyping algorithm to different datasets, the strength of our DA method will not change because it clearly shows that the different results will arise from the different characteristics of the study population, or the differences in available data or tools. Assessment of the impact of clinicians' diagnostic errors is outside the scope of this study and must be done in the future work. We aim to evaluate phenotyping algorithms developed using other techniques. For machine learning, one study indicated that the precision was almost unchanged but the recall differed according to the different training data [21]; thus, different results would be obtained from this study. We will report on which elements used in the DA method can be extracted automatically in the near future. This will lead to a systematic strategy for the development of phenotyping algorithms [12].

Conclusion

Our results confirmed that if researchers do not share annotation criteria in detail for classifying possible patients separately from definite CASEs and for classifying each type of possible patients, the characteristics of CASEs would differ among research teams; although phenotyping algorithms emphasize reproducibility, we cannot expect reproducible performance of the phenotyping algorithm. This was clinically rational for the four chronic diseases. In this study, we annotated EHRs using the DA method. This could increase reproducibility of retrospective EHR-based studies because it achieves annotations with low ambiguity by using information not directly referring to target diseases. We expect that our results will guide researchers on the reproducibility of EHR-based studies.

Ethics and acknowledgements

This research was approved by the Research Ethics Committee of the Graduate School of Medicine and Faculty of Medicine, The University of Tokyo (Permission number: 10733 (2015)). This work was supported by JSPS KAKENHI Grant Number 16J05555.

References

- [1] R. Woodfield *et al.*, Accuracy of electronic health record data for identifying stroke cases in large-scale epidemiological studies: a systematic review from UK biobank stroke outcomes group, *PLOS ONE* **10** (2015), e0140533
- [2] N. McCormick *et al.*, Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis, *PLOS ONE* **9** (2014), e104519
- [3] J.C. Kirby *et al.*, PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability, *JAMIA* **23** (2016), 1046-1052
- [4] C. Shivade *et al.*, A review of approaches to identifying patient phenotype cohorts using electronic health records, *JAMIA* **21** (2014), 221-230
- [5] R. Kagawa *et al.*, Development of type 2 diabetes mellitus phenotyping framework using expert knowledge and machine learning approach, *J Diabetes Sci Technol* (2016), doi: 10.1177/1932296816681584
- [6] R. Kagawa *et al.*, The technical problems for automated phenotyping, *Proc 30th JSAI* (2016), 1-4 (in Japanese)
- [7] D.M. Eddy, Variations in physician practice: the role of uncertainty, *Health Affairs* **2** (1984), 74-89
- [8] R.J. Carrol *et al.*, Naive electronic health record phenotype identification for rheumatoid arthritis, *AMIA Annu Symp Proc* **2011** (2011), 505-511
- [9] M.D. Ritchie *et al.*, Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record, *JAHG* (2010), 560-572
- [10] T. Lingren, Autism data dictionary. Available from: <https://phekb.org/sites/phenotype/autism> [cited 2016 Dec 9]
- [11] S. Bielinski, Heart failure with differentiation between reduced and preserved ejection fraction phenotype algorithm pseudo code cohort version. Available from: https://phekb.org/sites/phenotype/files/HF_algorithm_Cohort.pdf [cited 2016 Dec 9]
- [12] R. Richesson and M. Smerek, Electronic health records-based phenotyping. Available from: <https://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/> [cited 2016 Dec 9]
- [13] R. Duan *et al.*, An empirical study for impacts of measurement errors on EHR based association studies. *AMIA Annu Symp Proc* **2016** (2016), 1764-1773
- [14] U. Beuers *et al.*, Changing nomenclature for PBC: from 'cirrhosis' to 'cholangitis', *Hepatology* **62** (2015), 1620-1622
- [15] The Japan Diabetes Society, Evidence-based practice guideline for treatment of diabetes in Japan. Available from: http://www.jds.or.jp/modules/en/index.php?content_id=44 [cited 2016 Dec 9]
- [16] The Japanese Society of Hypertension, Guidelines for the management of hypertension 2014, 2014 (in Japanese)
- [17] Research Group on Intractable Liver or Biliary Tract Disease, Clinical guideline for primary biliary cirrhosis. Available from: http://www.nanbyou.or.jp/upload_files/PBC_guideline.pdf [cited 2016 Dec 9] (in Japanese)
- [18] Research Group on Idiopathic Hematopoietic Disorder, Reference guideline for clinical practice of autoimmune hemolytic anemia. Available from: http://zoketsushogaihan.com/file/guideline_H26/AIHA.pdf [cited 2016 Dec 9] (in Japanese)
- [19] A.N. Kho *et al.*, Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study, *JAMIA* **19** (2012), 212-218
- [20] A. Turchin *et al.*, DITTO – a tool for identification of patient cohorts from the text of clinician notes in the electronic medical record, *AMIA Annu Symp Proc* **2005** (2005), 744-748
- [21] R.J. Carrol *et al.*, Portability of an algorithm to identify rheumatoid arthritis in electronic health records, *JAMIA* **19** (2012), e162-e169
- [22] A. Karch *et al.*, Partial verification bias and incorporation bias affected accuracy estimates of diagnostic studies for biomarkers that were part of an existing composite gold standard, *J Clin Epidemiol* **78** (2016), 73-82

Address for correspondence

Rina Kagawa, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan
+81-3-5800-6427 kagawa-r@m.u-tokyo.ac.jp

Analysis of Historical Medical Phenomena Using Large N-Gram Corpora

Zdenko Kasáč^{a,b}, Stefan Schulz^a

^aInstitute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria,

^bFaculty of Medicine, Masaryk University, Brno, Czech Republic

Abstract

Historically, numerous indirect references to real world phenomena have been conserved in literature. High-quality libraries of digitized books and their derivatives (like the Google NGram Viewer) have proliferated. These tools simplify the visualization of trends in phrase usage within the collective memory of language groups. A straightforward interpretation of these frequency changes is, however, too simplistic to draw conclusions about the underlying reality because it is affected by several sources of bias. Although these resources have been studied in social sciences and psychology, there is still lack of user-friendly, yet rigorous methods for analysis of phenomena relevant for medicine. We present a methodological framework to study relationships of observable phenomena quantitatively over periods, which span over centuries. We discuss its suitability for knowledge extraction from current and future large-scale, book-derived, n-gram collections.

Keywords:

Historiography; Semantics; Publications

Introduction

It is a common in science for the phenomena of interest to be observed by proxy. Appropriate tools and methods are needed whenever direct measurement or inspection are unavailable or too complicated. A well-known example in the field of climatology is the analysis of ice core records for the estimation of atmospheric carbon dioxide [1]. Measuring the proportion of air gases trapped in polar ice allowed approximating parameters of atmospheric conditions along the last 800,000 years. Such information has redefined our understanding of the Earth's climate. In this case, nature did the scientist's job of archiving samples of air for hundreds of thousands of years, until our species developed and created a method to analyse the precious archived sample bank.

We seek to develop an analogous method for studying relevant phenomena for medicine and health. The described method would be also easy to exploit in other fields, e.g. in social sciences. A phenomenon might be anything (e.g. event, thing, situation, idea or sensation) that might be observed. On one hand, using the current epidemiological tools to research past phenomena would introduce recall bias and hide many environmental exposures because one simply forgets what happened many years ago. In other words, asking individuals about their expositions much earlier in their lives just would not work. For similar reasons, asking the respondents about the risk factor exposure of their great-grandparents would not work at all. It is therefore hard to analyse long-running health-related phenomena of the past directly. On the other hand, the extraction of retrospective data from medical records is overly limited, because the current form of medical documentation is recent, having started approaching its current form only in the run of the last century [2]. In addition, medical language has

constantly evolved, as did the concepts in physiology and pathology. Many aspects of health today considered important, and thus relevant for research, were granted low priority in earlier decades. Diagnostic criteria frequently change, and diagnostic tools evolve rapidly, which further complicates the evaluation of long-term epidemiological data. Finally, the availability of medical care constitutes an additional bias whenever studying medical records.

Therefore, to approximate and plot the temporal change in presence of a phenomenon, a source of coherent data is required. An ideal source must retain six properties of the ice core, viz. i) Populated with contemporary data; ii) Conserved unchanged until analysed; iii) Sufficiently large for an evaluation iv) Attributable to a certain population; v) The circumstances of data entry are homogeneous; vi) The process of data entry is self-organized, not institutional (to avoid group bias).

It seems that such a dataset exists. It is scattered across the world in the form of books in libraries. These books were written by authors who unwittingly – associatively (in fiction) or deliberately (non-fiction) reported on what they themselves saw/experienced (direct) or heard/read about (indirect observation) prior to writing. Once the books are digitised, their content can be reprocessed into databases of yearly n-gram frequencies [3]. By “n-gram”, we mean a space-delimited sequence of 1-grams. In this text, we use 1-gram interchangeably with token to denote a string of characters not separated by a space – mostly a word like “syphilis” or “antibiotic”. Regarding syphilis, Figure 1 depicts the frequency of its mention in literature, growing in the era of industrialization and peaking during wars, accompanied by its then most popular mercurial therapy. Both the disease and the useless therapy were slowly extinguished after the discovery of penicillin.

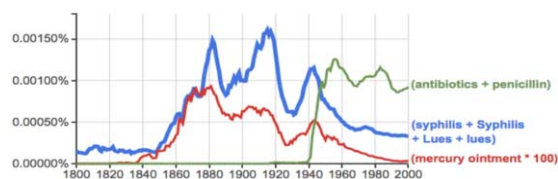


Figure 1 – The three plots, each sum of n-gram frequencies, from top: syphilis; traditional therapy; and effective therapy.

In the following, we propose and discuss a method to study real-world phenomena by systematic analysis of their mentions in literature over time. After cautious exclusion of potential confounders, we expect this mirroring in literature can be used for a quantitative visualization of the change in the real-world presence of certain phenomena – e.g. illnesses, observable factors of environment, etc. – as shown in the following.

Methods

Source

We used the freely available English (E), German (G) and Russian (R) Google Books n-gram data (V.20120701) as an input for used examples, documented in each caption as *Src:E/G/R*. Just the English version of the library used to create the Google n-grams contained 361 billion words derived from a nearly 5.2 million books large subset of Google Books library – approximately 4% of all books ever published [4]. The creators of Google NGram published descriptive statistics of 1-grams, 2-grams, 3-grams, 4-grams and 5-grams (text fragments, which contained 0-4 spaces) in these books. This resulting n-gram database consists of yearly counts of every n-gram and counts of books containing it. Our method requires precise metadata of the resource, in particular, precise publication date. Some kinds of publications notoriously contain older text, thus their publication date is imprecise. An example of these are periodicals and books that aggregate more works – such as anthologies or collected works by a given author. To improve the accuracy of the publication dates the corpus creators had removed these. After the filtering, a sample of book metadata had been verified by a human annotator (n=1000, sampling five books for every year between 1801 and 2000) and the rate of books with a publishing date outside a five-year range of tolerance was found to amount to 5.8% [4]. Though not required for the proposed method but justified by the propaedeutic character of this paper, we selected only few marker n-grams for the examples. In most cases, we averaged data over three years for our figures. If the span of moving average differed, we specified it in the caption (as $MA_{Avg=n}$), where n is the number of years. As explained later in this section the absolute ordinate value (n-gram frequency) is not relevant and is therefore usually relative, but always of linear scale starting from zero.

There is a broad topic spectrum of books in libraries and they do not come from a single group of authors. We appreciate this as such “institutional” origin could amplify group think bias [5]. There is yet another reason why we prefer books to news articles or abstracts of research papers: Due to Heaps’ Law [6], the longer the text (book vs. article), the higher the number of distinct n-grams in it. Similarly, also the number of topic-unrelated reports will follow the text length sub-linearly.



Figure 2 – “Radiogram” and “roentgenogram”. The ordinate is NOT scaled.

Arguably, trends in word usage are not merely due to phrases and words going in and out of fashion as seen in Figure 2. It is also, because their referents (the real-world *phenomena*) change in frequency and salience. Huge amounts of data have been generated by analyses of millions of digitized books [4]. The data has been applied in other fields, such as in language research [4, 7-8], social sciences such as measuring social functions and even predicting wars [9-10] and in psychology for measuring emotions, individualism or misery [11-13]. Yet it has remained virtually untouched by medical research.

One might wonder why this opportunity has been missed – it could have at least served as a cost-effective tool for the gen-

eration of population level hypotheses or for further observation of already known causal relationships.

There are objective reasons for this mistrust. Apart from its novelty, it may be explained by the uncertainty about whether the detected trends and correlations in word frequencies are really caused by the (i) changes in the domain of reference (real-world phenomena), or by (ii) linguistic and other confounders (Figure 2). We will discuss both in the following.

The division we just introduced is slightly different but not entirely unlike the division provided by Michel et al. [4] into “two central factors”: (i) cultural and (ii) linguistic change, such as the changing likelihood the author describes the referent by “Xray” vs. “radiogram” [8]. We had to define these groups differently, as many cultural factors – albeit important for others – confound or hide what we seek.

Let us shortly discuss the major linguistic origins of signal cancellation, noise or confounding. Aside from new words coming (*word birth*) to compete and eventually replace the current ones (which might face *word extinction*) just like seen in Figure 2, they might *acquire new spellings or meanings*, *lose* old referents and *undergo shift* in their meanings (see Figure 4). Interestingly, not only vocabulary gradually changes, so do parameters of the language dynamics. To provide an example, in Figure 3 we show a marked shortening of our *collective working memory* over the past two centuries. This supports the finding reached by other means by Petersen et al. [8]. Although we do not expect this to interfere with the method, it might still be important to keep this plot in mind when interpreting any trends in n-grams generated from that period.

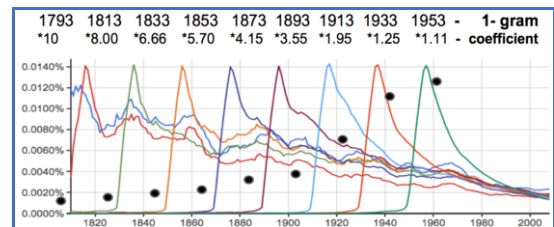


Figure 3 – Frequencies of 9 strings marking years starting with “1793” by increments of 20. A linear coefficient was used, so that the peak of each string would reach $f=0.014\%$. The black dots that follow each peak show inverse values of the coefficient used (roughly approximating the original amplitudes). The time for the recall to fall to 25% of the peak frequency notably shortens over the last 200 years. Apart from the “shortening of recall” a common “oscillation in recall” of these markers with a period of 20-25 years is visible. (Src:E).

In order to study the underlying real world phenomena, we should suppress the confounding linguistic effects. We can do this by appropriate selection of only high-precision *marker n-grams*, as described in the following section. Afterwards, we can approximate the environmental factor – the phenomenon.

A *phenomenon* in a domain is mirrored in literature by a *marker set* = a set of high-precision marker n-grams. E.g., phenomena of means of transportation in the real world are mirrored in books by a *marker set* $S_{\text{Means of Transportation}} = S_{\text{Mot}}$.

Marker set is not to be confused with the term *topic* used in n-gram analysis by others [14-15]. While a *topic* is defined by a group of n-grams that denote various phenomena and which typically occur together in a discourse on a theme, a *marker set* includes only the tokens that have a high precision (not just recall or sensitivity) in denoting the *phenomenon* of interest.

In our example, we assign a few case-insensitive strings to the example marker set for “Means of transportation” (S_{Mot}):

$$S_{\text{Means of transportation}} = S_{\text{MoT}} \ni \begin{cases} sS_{\text{Human-powered MoT}}, \\ sS_{\text{Animal-powered MoT}}, \\ sS_{\text{Motor-powered MoT}}, \\ \dots \text{etc.} \end{cases}$$

We used "...etc." as a wildcard for further possible *marker subsets* (sS_{xyz}) or n-grams, such as "travel" or "transport", which have good precision to S_{MoT} but do not fit in any of the *subsets*. In contrast, a token like "balloon" would not properly fit into S_{MoT} . It is imprecise for MoT because it could just as well denote a hot air balloon as an inflatable toy balloon.

N-grams belonging to a *marker set* can be assigned to categories called *subsets* (sS). We demonstrated this by dividing the *set* S_{MoT} into the following subsets and populating them:

$$\begin{aligned} sS_{\text{Human-powered MoT}} &= sS_{\text{H-MoT}} \ni \\ &\{ \text{"bicycle", "ride a bicycle", "skateboard", "roller"} \\ &\quad \text{"skates", "kayak", "by foot", "to walk" ... etc.} \} \\ sS_{\text{Animal-powered MoT}} &= sS_{\text{A-MoT}} \ni \\ &\{ \text{"dogcart", "horsecar", "riding horse",} \\ &\quad \text{"dog-drawn", "brancard", "howdah" ... etc.} \} \\ sS_{\text{Motor-powered MoT}} &= sS_{\text{M-MoT}} \ni \\ &\{ \text{"passenger car", "electric vehicle",} \\ &\quad \text{"aircraft", "airplane" ... etc.} \} \end{aligned}$$

Some of the n-grams truly belong to a *marker set* only during a certain era; e.g. more and more bicycles are now equipped with an electric motor. Therefore, if we are to define a subset: $sS_{\text{Exclusively Human-powered}}$, we would have to replace "bicycle" with a higher-precision synonym, or alternatively, limit the analysis to the time when "bicycle" meant solely an exclusively human-powered vehicle. For this reason, it is important to state the mutual exclusivity of the *marker subsets*.



Figure 4 – a) Left: "disease" has low and "illness" has high precision for human health problems. b) Right: The umbrella term "bronchoscope" strives not only during wars, but also with the arrival of a "flexible bronchoscope". Mentions of "rigid bronchoscope" were absent (1898-1958/66). (Src:E)

A typical example of an n-gram unfit for a *marker set* due to a non-intuitively low *marker precision* is "rigid bronchoscope", shown in Figure 4b. Since the introduction of bronchoscopy until 1960s, all bronchoscopes used to be rigid. In spite of this, or better to say: exactly for this reason, the term "rigid bronchoscope" did not need to exist. With introduction of fibre optics, flexible bronchoscopes became standard. To denote these flexible ones specifically, authors could now use terms "bronchofiberscope" and "flexible bronchoscope", but often, they would use the umbrella term. Due to this bandwidth steal, the "flexible bronchoscope" seems to be less frequent. Similarly, the name "First World War" (instead of "Great War") came to use only after advent of the Second World War – when a distinction was required. Such terms might decrease recall.

Now it is clear that for an n-gram to be added to a *set*, it must have a high *marker precision* – not to be confused with *meaning precision*, the mere absence of polysemy. *Marker precision* is a property of an n-gram that defines how well it reflects the phenomenon of interest across the whole period of analysis, by its frequency change in a given corpus, without detect-

ing unrelated phenomena. If only highly precise *marker* n-grams are present in the *set*, the *set* will mirror its phenomenon in a reliable way.

Non-precision might be also constant in time. This applies also for "disease" (see Figure 4a) which is astonishingly unspecific to human health, being also used for plants and animals. In contrast, the less popular "illness" shows both meaning and marker sensitivity to human medicine. We show another example in Figure 5, which highlights how cultural and political differences between the studied language groups have to be accounted for, like state censorship in the Soviet Union.

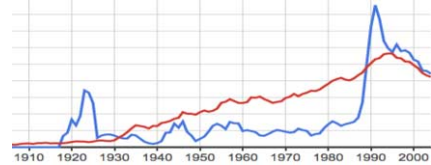


Figure 5 – The effects of politics on the literature [until 1917], first mentions during the February Revolution (blue, "свободный рынок", Src:R) and soaring interest in the following 8 years, immediately interrupted in the year of Lenin's death (1924). In red is the English "free market" (Src:E) which continued rising steadily until the mid '90s. (MAvg=1)

In the following paragraphs, we explain the steps by which noise and confounding effects can be minimised. As a result, the *marker sets* and *subsets* should be "clean" enough to relate quantitatively to the changes in the underlying phenomena.

Data pre-Processing

Creation of the "candidate" n-gram sets and subsets:

After selection of the phenomena for the observational analysis, we can define the future *marker sets* and *subsets* for these; e.g. as shown above: $[C_{\text{MoT}}, sC_{\text{H-MoT}}, sC_{\text{H-MoT}}$ and $sC_{\text{H-MoT}}$].

Mechanistic token selection [inclusion criteria]

Here we describe how to populate the main candidate sets of n-grams by *potential marker n-grams*. Then we propose how these could be cleaned from "noisy" n-grams by applying exclusion criteria on each of the "candidate" n-grams.

Firstly, we select the *potential marker* n-grams by:

1. extraction from defined and cited literature,
2. search of a comprehensive dictionary or thesaurus,
3. variation of the grammatical or lexical form,
4. a proposal from any the co-authors, validly argued by the rules 1-3. and/or validated through full-text analysis.

Even though *meaning precision* mostly implies *marker precision*, this is not always the case. As an alternative to the "mechanistic" selection of words described above, in some contexts it would be also viable to use token selection by questionnaire as done by others [13]. This approach is more useful for hard-to-define parameters or phenomena, such as emotions, attitudes or concerns because we cannot easily extract these from a dictionary. Even here, it must be reflected that like any other dimension of language these parameters are subject to change. If a word, such as "terrific" is charged with a positive emotion in 2016, this does not necessarily imply it was positive in 1916.

Definitions

In the next step, the timescale of analysis is defined and exclusion criteria are defined, which must filter out:

1. N-grams discovered that frequently denote unrelated phenomena (*homographs*) or have *low precision*.

- N-grams with a high noise to signal ratio, as with rare n-gram tokens or in older books due to OCR errors.

Where needed, the manual verification of precision for an n-gram might be performed by human annotator on a sample of full-text books – similar as performed by Michel et al. [13].

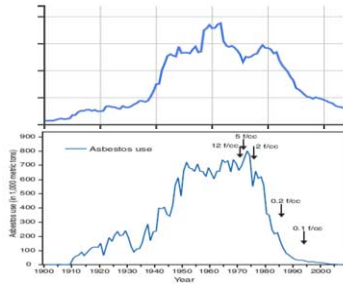


Figure 6 – (1900–2007): Upper plot shows a single bigram shape $C_c = \{\text{"asbestos cement"}\}$ (Src: E). Bottom plot shows asbestos use and legally permissible exposure limits as arrows in the USA [16].

Analysis

A measurement of n-gram frequency in a single point in time does not predict the contemporary expression of the related phenomenon in the real world by itself. It is also influenced by the salience of the phenomenon and the popularity of a given expression at that time. This introduces noise and makes a trend of the single marker n-gram hard to interpret. The shifts in popularity and the resulting noise can be suppressed by using the *marker sets*, thanks to the increased sample size.

As only the dynamics of the frequency of an n-gram tells the interesting story, the absolute amplitude of its trend is irrelevant for our analysis. We can linearly scale it without interference. In other words, the information is in the shape of curve, such as the positions of the local maxima, minima and trends. To extract knowledge, we propose the following approach.

For the first of the *marker sets* to be quantified (the *reference set*, usually the simplest one) and its subsets; if any, the plots of the member n-grams are added by simple summation.

Linear Fitting

Firstly, we plot the frequencies for each n-gram that belongs to the other *sets* (referred to as *independent sets*). We adjust their amplitudes by mutually independent coefficients, defined by a fitting algorithm. This algorithm uses the linear (!) coefficients with the goal to find the best alignment of the *independent set* to the *reference set* – measured by G (goodness-of-fit, see the section: Measurement).

If studying causalities with delayed effects (Figure 7), a proper adjustment of a whole (*sub-*) *sets* on the x-scale may be required as well, before proceeding with the following section.

Measurement

Now as we have fitted the curves of the *independent sets* S_i (or *sets*) to the *reference set* (S_r), we can measure two parameters. The first parameter is the **goodness-of-fit** ($0 < G < 1$). For a curve of each *set* S (or *subset* sS), the area under its curve A_c (or A_{sC}) is calculated. We define the goodness-of-fit as:

$$G = (A_{S_r \cap A_{S_i}}) / (A_{S_r \cup A_{S_i}})$$

An ideal correlation of two *marker sets* would have a goodness-of-fit $G=1$. Furthermore, a value of G for a set compared with a constant function (e.g. $x=1$) presents a calibration value $G=t$ for the given set. $G=t$ means “exactly no measurable association”. G lower than the calibration value is a negative correlation, higher G is a positive correlation.

The second parameter, called *share*, can be derived if G is close to 1. We define it as the fraction of the height of the *reference set* made up by a fitted *independent subset* in a time point. This correlates with the partial association of the respective sub-phenomenon (a category) in the respective time. Thus, we can use it to reflect a partial influence in case of a multifactorial causal relationship.

Results

In Figure Six, we compare asbestos use according to CDC with the frequency of mentions of asbestos cement in literature. The similarity of shape (growth in 1910, 1920s, and 1940) is still clear. Unlike the CDC data above, the mentions of “asbestos cement” are not limited to industrial and construction use of asbestos. The n-gram data can also report on demolitions, encounters in environment, daily use, etc. These would well explain the asynchrony in curve descent as even nowadays it is easy to come across aged asbestos containing materials. On the other hand, our single token did not account for other uses than reinforcement of concrete (such as braking systems, stove lining, etc.). Analysis using more n-grams would address this.



Figure 7 – Two simple marker sets in 2012 German n-gram data, using unweighted summation. $S_a = \{\text{"Eternit"}, \text{"Asbestzement"}\}$ and $S_m = \{\text{"Mesotheliom"}, \text{"Pleuresotheliom"}\}$.

Here comes up another aspect, viz. a possible amplification of “asbestos cement” by medical and legal literature over the last decades. The recognized causality causes a backwards flowing influence, where these terms do increasingly co-occur. Unlike the underlying aetiological connection, which is delayed, this amplification is immediate: when an author states that a mesothelioma was caused by exposure to asbestos cement. Two possible solutions are at hand: i) to use a resource that allows for excluding books containing both n-grams, thus suppressing the effect, or ii) to limit the inference to the “naïve” era, when the causality was not widely known. Not doing so would decrease sensitivity of the method and the goodness-of-fit.

To extend the example of asbestos, we used two simple sets S_a and S_m (see Figure 7) on a different population (German language group). Here, the patterns expected for this causality are obvious. There are three major peaks in frequency of the *marker set* S_a (1913, 1937, and 1996). For mesothelioma, there are two peaks (1962, 1978) followed by a rising trend filling the last 2 decades of this plot. German epidemiological data confirm this late growth [17]. More interestingly, the two peaks for the *marker set* S_m follow first two peaks of *marker set* S_a with a lag of 49 and 41 years, respectively. This is in accord with the generally accepted latency of incidence for this malignancy between 40 and 50 years [17–18]. We expect the next peak of S_m between 2036 and 2046.

Discussion

The precision of the method critically depends on the elimination of unrelated influences. Therefore, very popular but lexically ambiguous (homonymous) tokens need to be sacrificed (Figure 4a). Even when excluded, these n-grams might soar in

popularity at expense of used markers, which could lower the sensitivity. Thus, in the language where this happens, correlations of certain phenomena might stay hidden (false-negativity). There is certainly much space for improvement of the existing n-gram corpora, as pointed out by Pechenick et al. [19]. One of the good critiques raised, is that the dataset does not measure *true* popularity of ideas, but the frequency with which they are generated [19]. In our case, what is bug for culturomics becomes a feature: popular beliefs and fashions would only confound what we study, *viz.* the incidental observations by authors of books, whether related or unrelated to the topic or author popularity¹.

A second critique raised against the Google n-gram corpora, less relevant in our context, points out the unequal distribution of popular vs. scientific content underlying the English corpus [19]. For our goal, the genre is of little importance if the size of a sampling unit (length of book vs. article) is retained and attention to corpus homogeneity is paid when highly technical terms are used as marker n-grams for observations. An unequal genre/topic distribution could interfere with the sensitivity of this method. However, we doubt the validity of this criticism; namely, we doubt that there is indeed a growing fraction of scientific literature in the 2012 English n-gram resource. A simple search in English Fiction for the token “research” shows a steep and continuous rising trend for the entire 20th century. This suggests that the growth might not stem from a sampling error but rather from a similar culturomical process as shown in [10] – a social function shift in favour of research.

A more relevant critique points at the OCR errors, which are quite prevalent in the pre-1800 texts [4, 19]. This may be improved in the future by more reliable OCR techniques. In this case, which would enable digging far into the past, new evidence regarding lexical changes, social stigmatization, censorship etc. may be found.

Conclusions

We described the main issues in the extraction of information about observable real-world phenomena from changes of n-gram frequencies in large time-indexed n-gram corpora, with the focus on establishing a standardized method for exploitation of these data in, but not exclusively, the public health research. We plan to make a user-friendly tool based on this method available to other researchers and authors.

We demonstrated some of the possible applications on examples of use relevant for the public health and we have shown how already some very simple *marker sets* might well reflect the related real-world phenomenon in its parameters.

Acknowledgements

This work has been done as a part of Erasmus+ stay with a financial support of European Commission. We are thankful to Martin Komenda and Jiri Pavlacký for their valuable feedback. This work would not have been possible without support of Bronislava Kasáčová, Zdenko Medior Kasáč and Jana Petrášková.

References

- [1] D. Lüthi, M. Le Floch, B. Bereiter, T. Blunier, J.M. Barnola, U. Siegenthaler, D. Raynaud, J. Jouzel, H. Fischer, K. Kawamura, T.F. Stocker, High-resolution carbon dioxide concentration record 650, 000–800,000 years before present, *Nature* **453** (2008), 379–382
- [2] R.F. Gillum, From papyrus to the electronic tablet: A brief history of the clinical medical record with lessons for the digital age, *Am J Med* **126** (2013), 853–857
- [3] Y. Goldberg, J. Orwant, A dataset of syntactic-ngrams over time from a very large corpus of English books. *Second Joint Conference on Lexical and Computational Semantics* **1** (2013), 241–247
- [4] J.B. Michel, Y.K. Shen, A.P. Aiden, A. Veres, M.K. Gray, Google Books Team, J.P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M.A. Nowak, E. Lieberman Aiden, Quantitative analysis of culture using millions of digitized books, *Science* **331** (2011), 176–182
- [5] I.L. Janis, Groupthink: psychological studies of policy decisions and fiascoes, Houghton Mifflin, Boston, 1982
- [6] L. Lü, Z. Zhang, T. Zhou, Zipf's Law leads to heaps' aw: analyzing their relation in finite-size systems, *PLOS ONE* **5** (2010), e14139
- [7] L. Borin, D. Dubhashi, M. Forsberg, R. Johansson, D. Kikkinakis, P. Nugues, Mining semantics for culturomics: towards a knowledge-based approach, *ACM* (2013), 3–10
- [8] A.M. Petersen, J. Tenebaum, S. Havlin, H.E. Stanley, Statistical laws governing fluctuations in word use from word birth to word death, *Sci Rep* **2** (2012), 313
- [9] C. Besse, A. Bakhtiari, L. Lamontagne, Forecasting Conflicts using N-gram Models, (2012), 124–127
- [10] S. Roth, Fashionable functions: A Google Ngram view of trends in functional differentiation (1800–2000), *Int J Technol Hum Interaction* **10** (2014), 35–58
- [11] A. Acerbi, V. Lamos, P.R. Garnett, A. Bentley, The expression of emotions in 20th century books, *PLOS ONE* **8** (2013), e59030
- [12] R.A. Bentley, A. Acerbi, P. Ormerod, V. Lamos, Books average previous decade of economic misery, *PLOS ONE* **9** (2014), e83147
- [13] J.M. Twenge, W.K. Campbell, B. Gentile, Increases in individualistic words and phrases in American books, 1960–2008, *PLOS ONE* **7** (2012), e40181
- [14] D.M. Blei, Probabilistic topic models, *Communications of the ACM* **55** (2012), 77–84
- [15] D. Hall, D. Jurafsky, C.D. Manning, Studying the history of ideas using topic models, *ACL* (2008), 363–371
- [16] Centers for Disease Control and Prevention (CDC), Malignant mesothelioma mortality, *MMWR Morb Mortal Wkly Rep* **58** (2009), 393–396
- [17] V. Neumann, S. Loseke, D. Nowak, F.J.F. Herth, A. Tannapfel, Malignant pleural mesothelioma, *Dtsch Arztebl Int* **110** (2013), 319–326
- [18] S. Prazakova, P.S. Thomas, A. Sandrini, D.H. Yates, Asbestos and the lung in the 21st century: an update: Asbestos and the lung, *Clin Resp J* **8** (2014), 1–10
- [19] E.A. Pechenick, C.M. Danforth, P.S. Dodds, Characterizing the Google books corpus: Strong limits to inferences of socio-cultural and linguistic evolution, *PLOS ONE* **10** (2015), e0137041

Address for correspondence

Zdenko Kasáč, 375720@mail.muni.cz

¹ Maybe this text does not really belong to the “culturomics” field created by the authors of the NGram Viewer, as the culture is not our focus. Thus, a better fitting name for this approach would be “exponomics” as it examines the observable exposures of a population (the “exponome” of a population - as opposed to the exposome, which measures all the exposures of an individual).

An Incremental Adoption Pathway for Developing Precision Medicine Based Healthcare Infrastructure for Underserved Settings

Suranga N. Kasthurirathne^a, Paul G. Biondich^{b,c},
Burke W. Mamlin^{b,c}, Theresa A. Cullen^{b,c}, Shaun J. Grannis^{b,c}

^a School of Informatics and Computing, Indiana University, Indianapolis Indiana, USA, ^b Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA, ^c Indiana University School of Medicine, Indianapolis, Indiana, USA

Abstract

Recent focus on Precision medicine (PM) has led to a flurry of research activities across the developed world. But how can understaffed and underfunded health care systems in the US and elsewhere evolve to adapt PM to address pressing healthcare needs? We offer guidance on a wide range of sources of healthcare data / knowledge as well as other infrastructure / tools that could inform PM initiatives, and may serve as low hanging fruit easily adapted on the incremental pathway towards a PM based healthcare system. Using these resources and tools, we propose an incremental adoption pathway to inform implementers working in underserved communities around the world on how they should position themselves to gradually embrace the concepts of PM with minimal interruption to existing care delivery.

Keywords:

Precision Medicine; Vulnerable Populations; Health Information Systems.

Introduction

The need to offer providers a more complete picture of an individuals' health has led to the advent of precision medicine (PM), which is defined as an emerging approach to optimize clinical decision-making by taking individual variability in genes, environment and lifestyle into account [1]. By harnessing measurements from multiple modalities such as clinical and genomic evaluations, environmental exposures, behavioral patterns, and many others, we can develop a far more comprehensive view of the patient's health status and its trajectory over time.

Despite the recent emergence of the PM initiative, its core underlying principals have been known for decades [2], and manifested over the years in various forms such as personalized medicine [3] and the learning healthcare system [4]. Further, PM inspired activities such as the delivery of patient specific allergy alerts [5] and tailored immunization guidance [6] have been routinely used for many decades. A notable difference between these and more recent PM initiatives is the intensified focus on so-called "-omics", which includes the study of a body of information such as the genome, proteome, metabolome (metabolites), transcriptome (RNA transcripts), autoantibody profiles, etc., [7] made possible by the sequencing of the human genome [8] and rapid advances in high-throughput laboratory technologies and systems approaches across the fields of computer and biological sciences [9].

While -omics based therapeutics have gained widespread publicity and demonstrate significant potential [10], there

remain barriers to translating and broadly implementing these discoveries in the care delivery process. For example, -omics based PM therapeutics are expensive and resource-intensive activities that currently are ill-suited for large-scale clinical care [11,12] because breakthroughs in many -omics based therapeutics will not be available for use until research and discovery activities progress over the next 5 - 10 years [13].

By definition, the scope of PM is much larger than -omics. PM promises to provide a more complete picture of an individuals' health by capturing actionable information detailing individuals' health status and their socioeconomic context to inform provider decision-making. Thus, the core concepts of PM are relevant to modern medical care and beneficial for both developed and underserved settings. Examples of easily implementable low cost PM-based care that do not rely on -omics based data include the identification of high-risk patients based on socioeconomic factors and the detection of adverse drug events based on demographic and racial/ethnic factors.

While -omics based PM initiatives are typically cost intensive and, thus, currently unsuitable for widespread use, the overall aims of PM are relevant and invaluable for healthcare delivery. Thus, all healthcare systems, even those of the underserved world, stand to learn from this approach. Recent focus on PM has led to a flurry of research activities across the developed world. How can understaffed and underfunded health care systems in the US and elsewhere evolve to adapt PM to address pressing healthcare needs? Which PM based solutions are most suitable for use across underserved settings, if any? Also, which needs should be prioritized as implementers begin the thought process leading toward a PM based healthcare system?

We advocate that PM is heavily dependent on leveraging both novel and existing data and knowledge sources to ensure the accessibility and availability of information surrounding an individual and their environment. Thus, not all PM based activities may be cost and resource intensive. In this paper, we offer guidance regarding the wide range of sources of healthcare data and knowledge, as well as other infrastructure and tools that could inform PM initiatives. Such elements may serve as low hanging fruit easily and incrementally adopted on a pathway towards a PM based healthcare system. Using these resources and tools, we propose a progressive adoption pathway to inform implementers across the underserved world on how they should position themselves to gradually incorporate the principles of PM with minimal interruption to existing healthcare delivery.

Methods

We identified data and knowledge sources that could be

leveraged to support PM-based care delivery based on findings of the Precision Medicine initiative cohort program [14] and seminal publications [1,15,16]. However, these cannot operationalize PM based care without appropriate supporting systems and processes. Thus, we also identified a list of key infrastructural components that would enable the use of data and knowledge sources for delivering appropriate PM based care.

Next, based on our own expertise and these seminal publications, we sought to describe an incremental adoption pathway for underserved settings that seek to adopt PM initiatives by proposing when each of the aforementioned data, knowledge and infrastructure components should be integrated into existing healthcare systems. We considered five criteria to evaluate where each component fit into our proposed adoption pathway:

1. The availability of robust technical tools / platforms for ready use (Availability): *Are mature toolsets with demonstrated evidence of use available for adoption?*
2. The level of provider and user education / training required for adoption (Workforce capacity): *Does the tool/approach require specialist technical skills? How difficult is it to develop local capacity to manage and operate these systems?*
3. The practicality of integrating it into the existing health ecosystem and workflows (Integration to healthcare): *Can the component be feasibly integrated into existing/emerging infrastructure? Will it lead to disruptions, and does it require significant behavior changes for use?*
4. Cost burden/sustainability: *How sustainable is care delivery using these new components?*
5. Clinical impact: *Does robust evidence indicate this element meaningfully improves clinical outcomes?*

We adopted the ThoughtWorks Technology Radar (TTR) [17] to present our incremental PM adoption pathway. The TTR is a living document prepared by the ThoughtWorks Corporation, a leading software development company, to assess the risks and rewards of existing and nascent technologies, and the strategic importance of each of these for their organization at a specific point in time. TTR presents a concise overview of techniques, tools, platforms/languages and frameworks (four quadrants of a circle), and recommend their strategic importance to technology organizations by allocating each of these to one of four bins - adopt, trial, assess and hold (four concentric rings of a circle). Given its intended purpose, TTR has won widespread interest across the technology industry. Its ability to represent the changing technology landscape also makes it ideal for visualizing the evolving PM domain.

In adopting TTR for our own needs, we modified it as follows. Rather than the four quadrants presented in the TTR, we propose three slices (data, knowledge and infrastructure) that are more suitable to represent the needs of PM delivery. (b) Rather than the four rings presented in the TTR, we presented three: **Adopt** (highly significant components that must be adopted immediately), **Aspire** (components that should be focused on only after the adopt phase is completed) and **Hold** (components that should be held off for a future date where further advances are made). We selected these rings as they better represented the needs of a rapidly evolving healthcare system and stakeholder needs.

As our model is intended to represent a static snapshot the current state of PM-based research, and not serve as a 'living document', we will forgo TTR's 'new', 'moved' 'or no change' icons to indicate how the influence of each component has shifted over time.

Table 1 - Data types, examples and sources

Id	Data type	Examples	'Exemplar' or 'Typical' Sources
<i>Source 1: Clinical, demographic and behavioral data</i>			
1	Patient demographics and contact information	Birthdate, gender, race, sexuality, address, educational status, occupation	(1A) EHR systems; (1B) Patient registries (PR)
2	Behavioral and lifestyle measures	Physical activity levels, smoking/tobacco/alcohol use, assessment of other risk factors	(2A) EHR systems; (2B) Smart phones, wearable and home based devices
3	Sensor-based data	Cardiac rate and rhythm monitoring, respiratory rate, physical activity etc.	(3) Smart phones, wearable and home based devices
4	Structured clinical data	Medication, problem and diagnosis lists, vital signs, lab results, family health history measures etc.	(4A) EHR systems; (4B) Non -omics based Laboratory Information Management Systems (LIMS)
5	Unstructured clinical data	Narrative free text data/reports, EKG/EEG waveform data, radiology/medical imaging	(5A) EHR systems; (5B) Picture Archiving and Communication Systems (PACS)
6	-Omics data derived from bio-specimens	Genomics, proteomics, metabolites, cell-free DNA, single cell studies, infectious exposures, standard clinical chemistries, histopathology etc.	(6) External/ancillary Laboratory Information Systems that support -omics testing
<i>Source 2: Geospatial socio-economic, environmental and public health data</i>			
7	Socioeconomic data	Education, unemployment and crime rates, access to transportation, social services, health resources etc.	(7) Various state, private/not-for-profit, and advocacy organizations
8	Public health data	Health insurance coverage, disease rates, life expectancy, obesity rates etc.	(8) Various state and private/not-for-profit monitoring organizations
9	Healthcare resource data	Data on healthcare providers and facilities	(9A) Provider registries; (9B) Health facility registries
<i>Source 3: Other</i>			
10	Data obtained from social media	Behavioral data, over the counter medication purchases, food / drug consumption etc.	(10A) Facebook, twitter, other social media; (10B) Pharmacy billing data

Table 2 - Knowledge types, description and sources

Id	Knowledge type	Description	'Exemplar' or 'Typical' Sources
11	Drug-drug interactions	Information on how a drug affects the activity of another when administered together.	(11) Drug knowledge bases, such as Medscape and Drugbank.
12	Drug-gene interactions	Information on how a patient's genetic metabolic affects their ability to clear a drug.	(12) The Drug Gene Interaction Database (DGIdb) [18]; PharmGKB [19]
13	Allergen databases	Information on protein allergens and a patient's susceptibility	(13) Online databases such as allergome.org, allergen.org etc. [20]
14	Genomics knowledge bases	Information on human genes, genetic phenotypes and disorders, and relationships between phenotype and genotype.	(14) Tools such as Online Mendelian Inheritance in Man [®] (OMIM) [21], HGVBbaseG2P [22]
15	Public Health Genomics	The use of genomic discoveries to improve healthcare and disease prevention	(15) Public Health Genomics Knowledge Base [23]

Table 3 - Infrastructure resources

Id	Infrastructure component	Description
16	EHR systems	(16) An electronic version of patients' current and past medical condition that may include all key administrative clinical data relevant to a patients care under a particular provider, including demographics, progress notes, problems, medications, diagnosis etc.
17	Picture Archiving and Communication Systems (PACS)	(17) A medical imaging tool that provides storage and access to images from medical imaging technologies such as ultrasound (US), magnetic resonance (MR), Nuclear Medicine imaging, positron emission tomography (PET), endoscopy (ES), digital radiography (DR), radiography (CR) etc.
18	Laboratory Information Management Systems (LIMS) and tools	(18A) Traditional LIMS (18B) LIMS that support -omics based testing (18C) -Omics based testing equipment
19	Patient, provider and facility registries	(19A) Patient registries (19B) Provider / health worker registries (19C) Facility registries
20	Disease registries	(20) Registries used to track clinical care and outcomes of a defined patient population suffering from one or more chronic diseases such as diabetes, coronary artery disease, cancer or asthma.
21	Medical terminologies	(21) Various clinical terminologies for reporting diseases (ICD), SNOMED, laboratory data (LOINC) etc.
22	Messaging standards	(22) Various standards introduced to exchange electronic health data across multiple systems. Examples: the Health Level 7 (HL7) family of standards and the DICOM standard.
23	Health Information Exchange (HIE)	(23) A well rounded HIE would consist of many components listed above, such as point of care applications (EHR systems), various registries and a Shared Health Record (SHR) etc.
24	Biobank systems	(24) A biorepository that stores biological samples for use in laboratory based (genomics and traditional) testing.

Results

Various components and their adaption

We identified the following data components (Table 1), knowledge components (Table 2) and infrastructure (Table 3) as necessary for enabling PM based care systems. The information presented in Tables 1, 2 and 3 were fit into a PM technology radar (Figure 1) based on the five evaluation criteria presented above. The PM technology radar is comprised of three slices (data, knowledge and infrastructure) and three concentric circles (adopt, aspire and hold).

Description of the PM Technology radar

The Data Slice

As step 1 we recommend collecting various unstructured and structured clinical data from EHR, PACS and LIMS systems that may already co-exist within a healthcare facility. As step 2, we recommend investigating and integrating other healthcare data such as various registries and other geospatial information sources that may already exist outside the borders of the immediate healthcare facility. This facilitates the use of non-clinical socioeconomic and public health data that could be used to paint a better picture of patients' overall health. Step 3 consists of -omics data and

personal data sources such as social networks and smart phones, wearable and home based devices. We also note that some data types such as patient demographics and contact information may be obtained from multiple sources that belong to different stages. In such an event, we recommend that data sources are adopted incrementally over each stage, with patient demographics captured from EHR systems in stage 1, and gradually supplemented with registry data during stage 2.

The Knowledge Slice

We recommend adopting drug-drug interaction information as step 1. All other knowledge sources are based on -omics, and thus, should be held off until step 3.

The Infrastructure Slice

As step 1, we advocate adopting core health IT components such as EHR, PACS and LIMS systems, together with medical terminology/vocabularies and messaging standards for collecting clinical data. Step 2 consists of incorporating different health infrastructure components that together, form a robust health ecosystem; ideally a fully functional health information exchange. We recommend that -omics based infrastructure components be placed on hold, to be enabled together with genomic based data and knowledge sources after appropriate advanced are made in the field.

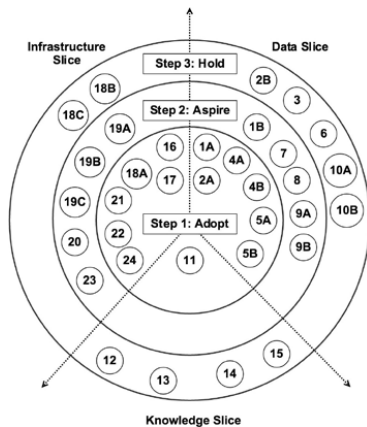


Figure 1 – PM Technology Radar

Discussion

Enabling PM based care systems can be seen as two distinct but related challenges: (a) optimizing the use of existing sources of readily available data and knowledge spread across clinical, socioeconomic, and public health spheres, and (b) the integration of -omics based therapeutics and other data obtained from smart devices and additional sources outside of the traditional healthcare delivery domain.

A fundamental requirement for all PM-based systems is the efficient management and use of structured information to inform providers' view of patients' health status. Thus, basic EHR systems that are widely used within underserved settings can serve as an initial stepping-stone on the incremental pathway toward building a more complete PM based healthcare system. To ensure comprehensive adoption of PM- based systems, readily obtainable clinical and behavioral data must also be sustainably captured and appropriate infrastructure components must be built to ensure that this data, together with additional sources of knowledge, are available for actionable use.

In evaluating the PM technology radar, it is evident that our selection of priorities for each step matches a specific pattern; the collection of data, knowledge, and infrastructure defined as step 1 (adopt) implies that the first step in enabling PM based care delivery lies in consolidating standardized data collection within a healthcare facility. As per figure 1, it is also evident that a significant portion of infrastructure resources should be implemented during the early stages of the process, as terminology and messaging standards are crucial to ensure standardized data collection.

The components assigned to step 2 (aspire) characterize expansion of data collection to resources that may lie beyond traditional disconnected healthcare facilities and towards a greater Health Information Exchange (HIE) spreading across a larger demographic area and resources. Adopting an HIE represents significant investment. However, many tools and platforms that enable HIE-based infrastructure have been in existence for years, and are available free of charge [24].

There is also significant value in shifting towards an HIE as these components can contribute to, and inform many other needs beyond PM based care delivery. In comparison, resources identified as step 3 (hold) represent a consolidation of -omics and other resources that have the potential to

significantly impact healthcare delivery, but are too immature and/or expensive for adoption at current time.

The PM technology radar also indicates that for underserved settings, early efforts should focus on using PM based initiatives to improve more common care delivery needs such as identifying patients at higher risk, better clinical decision support and medication adherence, as opposed to introducing new treatment processes for specific and less common diseases.

While strategies for adopting resources defined under step 1 are increasingly well understood, step 2 may not be as simple, and would require significant buy in and policy changes. Our approach includes significant emphasis on realizing a concordant HIE. However, this may be restricted by national or regional scale policies or law. For example, U.S. legislation bans creating a unique national patient identifier that could be used to effectively identify patients across the healthcare ecosystem.

Our reasons to delegate -omics based therapeutics as 'hold' for underserved settings are manifold, and based on practical, financial and policy based limitations rather than disagreement on the benefits of PM. Currently, genomic testing is available for approximately 2,000 clinical conditions in the U.S., and the number of available diagnostic tests is increasing exponentially [16]. However, genetics are only responsible for 20% of an individual's overall health status. Healthy behaviors (50%), environment (20%) and access to care (10%) pose significant impact on health, and are also relatively cheaper and easier to adopt [25]. Further, existing health information systems may not be geared to manage -omics based care. Research suggests that EHRs have poor support for online test ordering and provide limited decision support for genetic testing, interpretation of test results, and potential impact of results on patients and their families [26]. There is also significant need to develop, deploy, and adopt data standards to ensure data privacy, security, and integrity in managing PM based care delivery across these systems [27].

-Omics based treatment also raises significant questions regarding patient confidentiality and health payer rights and policies. In the U.S., these concerns led to the signing of the Genetic Information Nondiscrimination Act (GINA) [28]. However, other countries have been slow to follow. Given these considerations, and that projection that -omics based PM initiatives focused on discovery of disease risk factors, pharmacogenomics, disease biomarker discovery, loss-of-function mutations, new classification of diseases and clinical trials of targeted therapies are not expected to be realized over the next 5 to 10 years [13], we recommend that these components are placed on hold.

The use of smart devices for fitness and monitoring purposes is enjoying significantly greater interest and adoption in comparison to -omics based resources. However, we have assigned it to step 3 (hold) as; (a) smart devices do not contribute as significantly as other components and resources. These data often may also be collected via EHR systems or questionnaires, thus reducing their value; (b) despite tremendous interest, smart devices have not matured adequately to be robustly integrated with healthcare infrastructure, and may lead to security vulnerabilities. However, we acknowledge that smart devices present considerable potential, and may be moved to step 2 shortly.

Conclusions

The advent of PM represents the continuation of research

principals that have manifested in many forms over recent decades. Despite overemphasis on -omics based therapeutics, effective PM based care delivery also involves the efficient use of existing sources of clinical and public health data that are less expensive and easily integrated into existing healthcare infrastructure, and thus serve as 'low hanging fruits' for managing PM based care delivery. We present a PM technology radar that informs implementers on how they could gradually shift towards enabling better PM based care delivery.

References

- [1] F.S. Collins and H. Varmus, A new initiative on precision medicine, *New England Journal of Medicine* 372 (2015), 793-795.
- [2] National Research Council Committee on A Framework for Developing a New Taxonomy of Disease, *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease*, National Academies Press (US), 2011.
- [3] M.A. Hamburg and F.S. Collins, The path to personalized medicine, *New England Journal of Medicine* 363 (2010), 301-304.
- [4] A. Budrionis and J.G. Bellika, The Learning Healthcare System: Where are we now? A systematic review, *Journal of Biomedical Informatics* 64 (2016), 87-92.
- [5] S.A. Abookire, J.M. Teich, H. Sandige, M.D. Paterno, M.T. Martin, G.J. Kuperman, and D.W. Bates, Improving allergy alerting in a computerized physician order entry system, in: *Proceedings of the AMLA Symposium*, American Medical Informatics Association, 2000, p. 2.
- [6] P.R. Dexter, S. Perkins, J.M. Overhage, K. Maharry, R.B. Kohler, and C.J. McDonald, A computerized reminder system to increase the use of preventive care for hospitalized patients, *New England Journal of Medicine* 345 (2001), 965-970.
- [7] R.P. Horgan and L.C. Kenny, 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics, *The Obstetrician & Gynaecologist* 13 (2011), 189-195.
- [8] L. Nilsson, M. Hofflander, S. Eriksen, and C. Borg, The importance of interaction in the implementation of information technology in health care: a symbolic interactionism study on the meaning of accessibility, *Informatics for health & social care* 37 (2012), 277-290.
- [9] R. Chen and M. Snyder, Promise of personalized omics to precision medicine, *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 5 (2013), 73-82.
- [10] S. Morgan, J. Hurley, F. Miller, and M. Giacomini, Predictive genetic tests and health system costs, *Canadian Medical Association Journal* 168 (2003), 989-991.
- [11] A.S. Behl, K.A. Goddard, T.J. Flottemesch, D. Veenstra, R.T. Meenan, J.S. Lin, and M.V. Maciosek, Cost-effectiveness analysis of screening for KRAS and BRAF mutations in metastatic colorectal cancer, *Journal of the National Cancer Institute* 104 (2012), 1785-1795.
- [12] J.E. Dancy, P.L. Bedard, N. Onetto, and T.J. Hudson, The genetic basis for cancer treatment decisions, *Cell* 148 (2012), 409-420.
- [13] K. Hudson, R. Lifton, and B. Patrick-Lake, The precision medicine initiative cohort program—Building a Research Foundation for 21st Century Medicine, *Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director*, ed (2015).
- [14] K. Hudson, R. Lifton, and B. Patrick-Lake, The precision medicine initiative cohort program—building a research foundation for 21st century medicine, *Precision Medicine Initiative (PMI) Working Group Report to the Advisory Committee to the Director*, ed (2015).
- [15] S.J. Aronson and H.L. Rehm, Building the foundation for genomics in precision medicine, *Nature* 526 (2015), 336-342.
- [16] R. Mirnezami, J. Nicholson, and A. Darzi, Preparing for precision medicine, *New England Journal of Medicine* 366 (2012), 489-491.
- [17] ThoughtWorks Inc., TECHNOLOGY RADAR, in, 2016.
- [18] A.H. Wagner, A.C. Coffman, B.J. Ainscough, N.C. Spies, Z.L. Skidmore, K.M. Campbell, K. Krysiak, D. Pan, J.F. McMichael, and J.M. Eldred, DGIdb 2.0: mining clinically relevant drug-gene interactions, *Nucleic acids research* (2015), gkv1165.
- [19] C.F. Thom, T.E. Klein, and R.B. Altman, Pharmacogenomics and bioinformatics: PharmGKB, *Pharmacogenomics* 11 (2010), 501-505.
- [20] A. Mari, C. Rasi, P. Palazzo, and E. Scala, Allergen databases: current status and perspectives, *Current allergy and asthma reports* 9 (2009), 376-383.
- [21] J.S. Amberger, C.A. Bocchini, F. Schiattacate, A.F. Scott, and A. Hamosh, OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders, *Nucleic acids research* 43 (2015), D789-D798.
- [22] G.A. Thorisson, O. Lancaster, R.C. Free, R.K. Hastings, P. Sarmah, D. Dash, S.K. Brahmachari, and A.J. Brookes, HGVDbaseG2P: a central genetic association database, *Nucleic acids research* 37 (2009), D797-D802.
- [23] W. Yu, M. Gwinn, W.D. Dotson, R.F. Green, M. Clyne, A. Wulf, S. Bowen, K. Kolor, and M.J. Khoury, A knowledge base for tracking the impact of genomics on population health, *Genetics in Medicine* (2016).
- [24] R. Crichton, D. Moodley, A. Pillay, and C.J. Seebregts, An interoperability architecture for the health information exchange in Rwanda, (2012).
- [25] Bipartisan Policy Center, What Makes Us Healthy vs. What We Spend on Being Healthy, in, 2012.
- [26] M.T. Scheuner, H. de Vries, B. Kim, R.C. Meili, S.H. Olmstead, and S. Teleki, Are electronic health records ready for genomic medicine?, *Genetics in Medicine* 11 (2009), 510-517.
- [27] J.D. Tenenbaum, P. Avillach, M. Benham-Hutchins, M.K. Breitenstein, E.L. Crowgey, M.A. Hoffman, X. Jiang, S. Madhavan, J.E. Mattison, and R. Nagarajan, An informatics research agenda to support precision medicine: seven key areas, *Journal of the American Medical Informatics Association* (2016), ocv213.
- [28] K.L. Hudson, M. Holohan, and F.S. Collins, Keeping pace with the times—the Genetic Information Nondiscrimination Act of 2008, *New England Journal of Medicine* 358 (2008), 2661-2663.

Address for correspondence

Suranga N. Kasthurirathne,
School of Informatics and Computing (SOIC)
Indiana University, Indianapolis, IN, USA
Email: snkasthu@iupui.edu

Diagnostic Machine Learning Models for Acute Abdominal Pain: Towards an e-Learning Tool for Medical Students

Piyapong Khumrin^a, Anna Ryan^b, Terry Judd^b, Karin Verspoor^a

^a Dept of Computing and Information Systems, School of Engineering, University of Melbourne, Melbourne, Australia,

^b Dept of Medical Education, Melbourne Medical School, University of Melbourne, Melbourne, Australia

Abstract

Computer-aided learning systems (e-learning systems) can help medical students gain more experience with diagnostic reasoning and decision making. Within this context, providing feedback that matches students' needs (i.e. personalised feedback) is both critical and challenging. In this paper, we describe the development of a machine learning model to support medical students' diagnostic decisions. Machine learning models were trained on 208 clinical cases presenting with abdominal pain, to predict five diagnoses. We assessed which of these models are likely to be most effective for use in an e-learning tool that allows students to interact with a virtual patient. The broader goal is to utilise these models to generate personalised feedback based on the specific patient information requested by students and their active diagnostic hypotheses.

Keywords:

Decision Support Systems, Clinical; Formative Feedback; Artificial Intelligence

Introduction

Diagnostic reasoning is the cognitive process of deriving a correct diagnosis from a patient's presenting clinical problem. The development of diagnostic reasoning skills in medical training starts with a disease-oriented approach by learning common presentations of different diseases [1]. Given this knowledge, students approach patients with a presenting problem and hypothesise the most likely diagnosis. They gather patient information through the steps of history taking, physical examination, and consideration of the results of laboratory procedures and other investigations. Students accumulate the information to prune and prioritise possible diagnoses until they get to a final diagnosis [2]. During medical training, students practice diagnostic reasoning skills with patients under expert supervision, called "bedside teaching". Students gather patient information and present their diagnostic reasoning to the expert. The expert identifies errors, misconceptions and inadequacies and formulates suitable feedback to help students to reconstruct their knowledge [3]. Personalised feedback is a key element of learning and instruction [4-6] and bedside teaching is known to improve diagnostic reasoning skills. However, its use is declining due to a range of factors, including increased patient turnover, concerns with patient privacy, increased technology in the diagnostic process, increased numbers of students, and limited availability of experts [7; 8].

E-learning systems can help to address some of these challenges and when used effectively, in conjunction with traditional approaches, can aid in the development of diagnostic reasoning skills [9-11]. However, tailoring the

feedback from e-learning systems so that it is both effective and fits the needs of individual students is difficult [4; 12]. Effective feedback should help the student to identify what they already know or have mastered, where potential knowledge gaps or misconceptions lie, provide an indication of their learning progress, and support them to achieve their learning goals [13].

Clinical Decision Support Systems (CDSS) are computer systems that assist doctors to make decisions and are typically used in either the diagnostic process or to support clinical management. CDSS synthesise information based on patient data and use the information to generate a prediction. Prior research has demonstrated that using a CDSS to assist physicians' diagnostic and treatment processes can improve both the effectiveness and efficiency of patient care [14-16], a clear example being the management of acute abdominal pain [17]. Applying the concept of a CDSS, we aim to use machine learning models to produce personalised feedback within an adaptive game-based learning tool intended to support the development of medical students' diagnostic reasoning and decision-making skills. The learning tool will allow students to interact with a virtual patient, and revealing relevant patient information and diagnostic predictions in response to the students' actions and requests.

This paper presents the development of the machine learning model that underpins the e-learning tool. It describes the collection and processing of a large corpus of patient data, and the development, training, testing and comparison of alternative machine learning models based on these data. A preferred model is identified and the rationale for its selection in the context of the e-learning tool explained. We also discuss how this model may be leveraged within the tool to generate personalised and appropriate feedback for medical students.

Methods

Phase I: Data collection

Electronic patient records were collected to train the machine learning models. We captured and processed de-identified medical data from three disparate but complementary sources; 1) Student log cases 2) Student entered Electronic Health Records (created by medical students at our university as part of their course and stored within their curriculum delivery system 3) Electronic health records within the public hospital affiliated with our university's medical school. The data were used to develop clinical scenarios to be presented within the learning tool and to train machine learning models for of the learning tool's embedded CDSS. All data collection was approved either by our university's Health Sciences Human Ethics Sub-Committee or by the affiliated hospital's Human Research Ethics and Research Committee.

Phase II: Case selection

We selected cases from the electronic systems with the following inclusion and exclusion criteria:

Inclusion criteria

1. A principal diagnosis of one of five key conditions (appendicitis, gastroenteritis, urinary tract infection, ectopic pregnancy, or pelvic inflammatory disease).
2. The treatment protocol for the principal diagnosis was completed.

Exclusion criteria

1. More than one of the conditions of interest were diagnosed in an individual patient in the same admission.
2. A previous history of other conditions or procedures or treatments that would rule out one or more of the key diagnoses (e.g. appendectomy for appendicitis).

Two hundred and ninety-eight clinical cases were identified by applying the query criteria. Ninety cases were excluded leaving 208 valid clinical cases (see Table 1).

Phase III: Data pre-processing

PK manually extracted data from the selected cases and transformed the information into an array of features. Seventy-five features were extracted from history taking (n=48), physical examinations (n=13), laboratory and investigation results (n=13), and a target class (n=1).

Table 1 – Sample size per diagnosis

Diagnoses	n
Appendicitis (AP)	51
Gastroenteritis (GE)	53
Urinary tract infection (UTI)	68
Ectopic pregnancy (EP)	11
Pelvic inflammatory disease (PID)	25
Total	208

Phase IV: Machine learning training

We used pre-processed clinical features from Phase III to train machine learning models for classifying the features into one of the five target diseases. We utilised Weka version 3.8 [18] for the training process using 10-fold cross validation on a training set to evaluate alternative algorithms: testing Naïve Bayes, Support Vector Machine (SVM), Neural Networks (NN), C4.5 decision tree (J48), and Logitboost (using DecisionStump as a classifier). We used the correlation attribute evaluation to rank the level of feature relevance to predict a diagnosis. We intend to use one or more of the machine learning models to predict diseases when students raise possible diagnoses on the basis of clinical observations. We plan to transform the prediction of the classifier to a suitable form of feedback to represent the likelihood of diagnosis based on present findings. The rank of a given feature's relevance in the model will be used to guide students' feature selection.

We grouped the target classes in two ways. First, we treated the entire dataset as a single group and targeted classification of the five diseases considering all diagnoses together, which is a "multi-class classification" scenario. Second, we divided the target classes into two groups, to create a "binary classification". In this case, the first group contains one of the target diseases, with all others being assigned to the second group. The diagnoses in the second group are then merged into a single class – e.g. the classifier makes a decision between "appendicitis" and "not appendicitis".

Results

Multi-class classification result

Table 2 shows the percentage of overall correctly classified instances from six machine learning models. ZeroR (majority class classifier) provides a baseline performance. All classifiers predict better than the baseline but the top three classifiers were Logitboost, NaïveBayes, and Neural Network, shown in bold.

Table 2 – Accuracy of multi-class classification

Classifiers	% Accuracy
ZeroR	32.69
J48	66.35
SVM	72.60
NN	84.62
NaïveBayes	85.10
Logitboost	94.71

Table 3 shows the F1-measure of classification in different diagnoses. Logitboost predicts all diagnoses with the highest performance (F1-measure between 90 – 99%), and performed substantially better than all other classifiers in the case of EP.

Table 3 – F1-measure

	AP	GE	UTI	EP	PID
ZeroR	0.000	0.000	0.493	0.000	0.000
NaïveBayes	0.887	0.857	0.889	0.571	0.793
SVM	0.755	0.755	0.786	0.455	0.480
NN	0.857	0.862	0.949	0.375	0.679
J48	0.706	0.627	0.744	0.167	0.500
Logitboost	0.923	0.925	0.993	0.900	0.939

The distribution of correct and incorrect classifications (confusion matrix) in different diagnoses appears in Table 4 for Naïve Bayes and Table 5 for Logitboost. Diseases in rows and columns represent the true and predicted diagnoses, respectively. AP, EP, and PID have the most instances of misclassification.

Table 4 – NaïveBayes confusion matrix

	predicted				
true	AP	GE	UTI	EP	PID
AP	47	1	2	1	0
GE	4	45	0	0	4
UTI	1	6	56	2	3
EP	2	0	0	6	3
PID	1	0	0	1	23

Table 5 – Logitboost confusion matrix

	predicted				
true	AP	GE	UTI	EP	PID
AP	48	3	0	0	0
GE	3	49	1	0	0
UTI	0	0	68	0	0
EP	1	0	0	9	1
PID	1	1	0	0	23

Table 6 shows selected key decision features from J48 decision tree (not shown) which are correlated to clinical knowledge.

Table 6 – Key decisions for the diagnoses on J48 decision tree

Diagnoses	Features
AP	Right lower abdominal pain
GE	Upper abdominal pain Diarrhea
UTI	Dysuria Lower abdominal pain
EP	Serum hCG
PID	Serum hCG

Binary classification result

Table 7 shows the F1-measure score of binary classifications for individual diagnoses, using the NaïveBayes and LogitBoost classifiers. The first and second sub-columns under the classifier column represent the first and second groups in the binary classification, respectively. The last sub-column is the average F1-measure in the first two sub-columns. LogitBoost classifies all target diagnoses with good accuracy whereas NaïveBayes registers a significant drop for EP.

Table 7 – F1-measure of the binary classifications

	Naïve Bayes			LogitBoost		
	X	not X	Avg	X	not X	Avg
AP	0.851	0.952	0.928	0.832	0.949	0.918
GE	0.874	0.958	0.937	0.882	0.962	0.942
UTI	0.894	0.951	0.932	0.963	0.982	0.976
EP	0.500	0.969	0.945	0.952	0.997	0.995
PID	0.724	0.955	0.928	0.939	0.992	0.985

Feature selection

Feature selection on multi-class classification

The most highly ranked features with respect to the multi-class classification were a history of left abdominal pain and the patient's age. The top ten features, all of which returned a percentage relevance of at least 15% are listed in Table 8.

Table 8 – Top 10 features of the multi-class classification

Ranked features	% relevance
Left abdominal pain history	31
Age	25
Lower abdominal tenderness	20
Guarding	19
Upper abdominal pain history	18
Vomiting	16
Rovsing's sign	16
Rebound tenderness	16
Leukocyte in UA	15
Pain quality	15

Feature selection on binary classification

We also measured the correlation of features to individual diagnoses, based on the binary classification model. Table 9 shows selected features across the five diseases. Symptoms such as age, gender, location of abdominal pain, characteristics of pain, fever, gastrointestinal, urinary tract system, and gynaecological histories are useful in distinguishing between the different diseases. Right lower abdominal pain and pain migration are more specific to AP. Diarrhea is more specific to GE. Urinary tract symptoms are more specific to UTI. UPT and serum hCG are most specific to EP, while gynaecological symptoms and laboratory result

for sexual transmitted diseases are most strongly correlated with PID.

Table 9 – Selected features of the binary classifications

Features	Relevant features
AP	Right lower abdominal pain, pain migration
GE	Diarrhea
UTI	Dysuria, urine colour, haematuria
EP	UPT, serum HCG
PID	Vaginal discharge, PCR for Chlamydia

Discussion

Retrospective vs prospective data collections:

We used a retrospective data collection method to collect clinical cases for training machine learning models. Unlike prospective approaches, this meant we were unable to constrain the records within a single template. Further limitations of this method included the variable formatting and structure of records, missing values, and temporal data. Our reliance on different data sources (even within the one hospital it is not unusual for different clinical departments to use different electronic health record systems) created a range of data entry and processing issues related to feature definition, type and sequence as well as clinical interpretation. This variation introduced considerable noise, reducing data validity and complicating data pre-processing. Many of these issues would undoubtedly have been reduced in the case of prospective data collection, as requirements or recommendations around record creation can be more closely defined and monitored. However, prospective data collection would have required considerable additional planning on our part, and carried substantial additional time and administrative costs that are likely to have been unacceptable to hospital staff given their heavy workloads.

Table 10 shows pros and cons of retrospective and prospective data collection. In summary, retrospective data collection is simpler with lower costs and does not impact on patients' treatment. Prospective data collection, on the other hand, provides higher quality data and data validity but is likely to involve unacceptable costs.

Table 10 – Pros and cons between retrospective and prospective data collection

Factors	Retrospective	Prospective
Budget	less	more
Increase work for treatment process	no	yes
May influence treatment process	no	yes
Sample size	flexible	restricted
Quality of data	poor to good	very good
Data validity	poor to good	very good

For this study, we utilised case records provided in three different digital formats: plain text, scanned documents (images), and Word documents. The majority of these were extracted from the hospital's record databases. The number of records available to us were sufficient for our purposes except in the cases of EP and PID, because patients diagnosed with those two diseases tend to be admitted to a specialist women's hospital rather than the general hospital involved in our study. Inclusion and exclusion criteria were used to filter the cases because certain histories have a critical impact on formulating a list of possible diagnoses. For example, a patient who has had their appendix or ovaries removed should never be

diagnosed with AP or EP, respectively. Those histories provide a spot conclusion rather than enhancing the development of diagnostic reasoning processes.

Data pre-processing

Before PK started extracting data, he listed common presentations of the five diseases from medical standard textbooks [19-23]. He used standard medical terms from UMLS [24] and SNOMED CT [25] to identify and organise features and synonyms. Seventy-two (72) of 75 available features contained some missing values (only age, gender, and target diagnosis were complete). Missing values commonly occur because admitting doctors deem it as self-evident, unnecessary or irrelevant. For example, doctors will never ask questions about menstrual history, or order a urine pregnancy test and serum hCG for male patients. Similarly, a normal urine finding typically infers a negative finding in relation to all abnormalities of the urinary tract.

Symptoms also typically change over time. For example, the location, severity and quality of pain, and whether the pain is relieved by medication, can all change as a disease progresses. In the case of the location of pain, we divided pain location into two episodes – prior to and at admission. If these were different then the pain had migrated. More generally, we only used symptoms and signs from records created in the emergency department or on first admission to a hospital ward to reduce the effects of timing and treatment.

Clinical interpretation of classification results

The top three classifiers were Logitboost, NaïveBayes, and Neural Network, which all had a predictive accuracy above 80% (see Table 2). The Logitboost algorithm improves the results of a classification by reweighting mis-classified samples and taking a weighted majority vote to form training data [26]. It returned a significantly higher overall predictive accuracy than either NaïveBayes or Neural Network classifiers due to its superior predictive performance for EP (see Table 3). The classification performance for EP was low for all other algorithms due to a combination of low record numbers and missing data. We believe that Logitboost performed well despite these issues because of its use of a Decision Stump; a subroutine within Logitboost that analyses patterns of missing data to develop rules for classification.

When we considered the distribution of misclassified EP over other diseases predicted by NaïveBayes (Table 4) and Logitboost (Table 5), the models were more likely to mis-classify EP as either AP and PID, which have a number of overlapping symptoms, rather than UTI or GE. By way of comparison, in a series of early studies, de Dombal and colleagues [27; 28] used a CDSS employing a Bayesian classifier to predict abdominal pain within 600 prospectively collected clinical cases. They reported an overall diagnostic accuracy of 91.8%.

Table 11 compares the research methods and classification results for the current and these earlier studies when using a similar classification approach [27]. In this case, the machine learning model developed by de Dombal had the higher overall classification accuracy. We noted the key success factors in the de Dombal study were sufficient samples, equal distribution of sample, and the quality of input data.

While all classifiers apart from the baseline (ZeroR) provided more accurate predictions than the J48 algorithm, it is the only one that produces human-readable output. Its decision tree style output is easily interpreted and has the potential to provide useful feedback to users during the decision-making process.

Table 11 – Comparison of research methods and classification result between two studies

Methods/Result	this study	de Dombal
Data collection	retrospective	prospective
Sample size	208	600
Machine learning method	NaïveBayes	NaïveBayes
Target classes	5	7
Overall accuracy	85.10	91.80

Key features

For multi-class classification, left abdominal pain appears to be a key feature as it positively selects for those diseases presenting with lower abdominal pain – appendicitis, ectopic pregnancy, and pelvic inflammatory disease. Gastroenteritis and urinary tract infection are less likely to be selected because the pain position for these two diseases is more diffuse. However, a history of left abdominal pain reduces the probability of appendicitis and indicates more strongly ectopic pregnancy or pelvic inflammatory disease. Age is also an important factor, with appendicitis, ectopic pregnancy and pelvic inflammatory disease most strongly associated with younger patients. In the case of the binary classification, we were able to identify key diagnostic features associated with each disease. We will use this information to help medical students to identify strongly relevant clinical information to support their diagnostic decisions.

Machine learning model selection

Our final choice of machine learning model for use in the proposed learning tool is informed by two requirements: clinically appropriate predictions; and classification performance. We give precedence to clinical interpretation because proper development of the diagnostic reasoning process is more important than maximising the number of correct decisions. Binary classification is preferred because it provides a better sense of scaling of the likelihood level than multi-class classification. And, while the overall prediction performance of NaïveBayes is slightly inferior to Logitboost, the NaïveBayes predictions are more reflective of actual clinical judgements. For example, Logitboost predicts zero probability of ectopic pregnancy on a female patient presenting with right lower abdominal pain, which is clinically inappropriate. Accordingly, we selected the NaïveBayes classifier in combination with binary classification as our preferred model.

Feedback representation

We plan to use the machine learning model to provide two types of feedback (interim and final) within the proposed learning tool. Interim feedback will be based on the interpretation of predictive correlations between selected patient information and top three most likely diagnoses following the history taking, physical examination, and laboratory and investigations steps. Final feedback will present the correct diagnosis, and generate a user score based on how often the correct diagnosis is selected in the differential diagnosis list, correlation to the correct diagnosis, and the inclusion of key patient information.

We plan to use an overlap model to assess students' performance (and potential learning gains) while using the tool [29]. This treats the student's decisions (student model) as an incomplete model which can be compared to the machine learning model (complete model). The more similar these two models are, the higher the assessment of the student's performance.

Conclusion

Timely and clinically appropriate personalised feedback is key to the development of students' diagnostic reasoning skills. E-learning has a role to play here, through the provision of personalised and appropriately scaffolded feedback on students' diagnostic decision-making on virtual patient cases. We propose to use diagnostic models derived through machine learning as the basis for giving relevant feedback. In this paper, we describe how we trained machine learning models using a large corpus of real clinical cases to develop differential diagnoses related to presentations of abdominal pain. We selected a model that combines a Naïve Bayes classifier with binary classification for further use in the learning tool based on a combination of its predictive performance and the clinical relevance of that model's predictions.

Acknowledgements

PK thanks Chiang Mai University for a scholarship to support his PhD study. We also thank the Tilley family, Dr Chris Leung, and Professor Richard O'Brien of Austin Hospital Clinical School, the University of Melbourne for research support.

References

- [1] R. Patel, J. Sandars, and S. Carr, Clinical diagnostic decision-making in real life contexts: A trans-theoretical approach for teaching: AMEE Guide No. 95, *Med Teach* **37** (2015), 211-227.
- [2] A. Venot, A. Burgun, and C. Quantin, *Medical informatics, e-Health : fundamentals and applications*, 2014.
- [3] M. Garout, A. Nuqali, A. Alhazmi, and H. Almoallim, Bedside teaching: an underutilized tool in medical education, *Int J Med Educ* **7** (2016), 261-262.
- [4] M. Frize and C. Frasson, Decision-support and intelligent tutoring systems in medical education, *Clin Invest Med* **23** (2000), 266-269.
- [5] W.C. McGaghie, S.B. Issenberg, E.R. Petrusa, and R.J. Scalese, A critical review of simulation-based medical education research: 2003-2009, *Med Educ* **44** (2010), 50-63.
- [6] D.H. Jonassen, Assoc. Educational Communications and Technology, *Handbook of research for educational communications and technology: a project of the Association for Educational Communications and Technology*, Macmillan Library Reference USA, New York, 1996.
- [7] M. Peters, O. Ten Cate, Bedside teaching in medical education: a lit review, *Perspect Med Educ* **3** (2014), 76-88.
- [8] A. Salam, H.H. Siraj, N. Mohamad, S. Das, Y. Rabeya, Bedside teaching in undergraduate medical education: issues, strategies, and new models for better preparation of new generation doctors, *Iran J Med Sci* **36** (2011), 1-6.
- [9] P. Devitt and E. Palmer, Computers in medical education 1: evaluation of a problem-orientated learning package, *Aust N Z J Surg* **68** (1998), 284-287.
- [10] J.N. Hudson, Computer-aided learning in the real world of medical education: does the quality of interaction with the computer affect student learning?, *Med Educ* **38** (2004), 887-895.
- [11] J.G. Ruiz, M.J. Mintzer, R.M. Leipzig, The impact of E-learning in medical education, *Acad Med* **81** (2006), 207-212.
- [12] E. Vasilyeva, Towards personalized feedback in educational computer games for children, in: *Proceedings of the sixth conference on IASTED International Conference Web-Based Education - Volume 2*, ACTA Press, Chamonix, France, 2007, pp. 597-602.
- [13] B.J. Mason, R. Bruning, Providing feedback in Computer-Based Instruction: What the research tells us, 2001.
- [14] K. Kawamoto, C.A. Houlihan, E.A. Balas, and D.F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, *BMJ* **330** (2005), 765.
- [15] A.X. Garg, N.K. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam, and R.B. Haynes, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review, *JAMA* **293** (2005), 1223-1238.
- [16] R. Kaushal, K.G. Shojania, and D.W. Bates, Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review, *Arch Intern Med* **163** (2003), 1409-1416.
- [17] J.G. Cooper, R.M. West, S.E. Clamp, and T.B. Hassan, Does computer-aided clinical decision support improve the management of acute abdominal pain? A systematic review, *Emerg Med J* **28** (2011), 553-557.
- [18] E. Frank, I.H. Witten, and M.A. Hall, *Data mining : practical machine learning tools and techniques*, Morgan Kaufmann, Burlington, Mass., 2016.
- [19] N.J. Talley, S. O'Connor, A summary of the gastrointestinal examination and extending the gastrointestinal examination, in: *Clinical examination: a systematic guide to physical diagnosis*, CL Elsevier, Sydney, 2010.
- [20] N.J. Talley and S. O'Connor, Correlation of physical signs and gastrointestinal disease, in: *Clinical examination : a systematic guide to physical diagnosis*, Churchill Livingstone Elsevier, Sydney, 2010.
- [21] N.J. Talley, S. O'Connor, The gastrointestinal history, in: *Clinical examination: a systematic guide to physical diagnosis*, CL Elsevier, Sydney, 2010.
- [22] N.J. Talley, S. O'Connor, The gastrointestinal physical examination, in: *Clinical examination: a systematic guide to physical diagnosis*, CL Elsevier, Sydney, 2010.
- [23] J. Hall, Essentials of clinical examination handbook, in, Thieme, New York, 2013, p. 640 p.
- [24] D.A. Lindberg, B.L. Humphreys, and A.T. McCray, The Unified Medical Language System, *Methods Inf Med* **32** (1993), 281-291.
- [25] K. Bardadin, [SNOMed--international system for coding changes in medicine], *Pol J Pathol* **46** (1995), 203-205.
- [26] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, (2000), 337-407.
- [27] J.R. Staniland, J. Ditchburn, and F.T. De Dombal, Clinical presentation of acute abdomen: study of 600 patients, *Br Med J* **3** (1972), 393-398.
- [28] J.C. Horrocks, A.P. McCann, J.R. Staniland, D.J. Leaper, and F.T. De Dombal, Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases, *Br Med J* **2** (1972), 5-9.
- [29] J.L. Stansfield, *Wumpus Advisor I. A First Implementation of a Program That Tutors Logical and Probabilistic Reasoning Skills*. Distributed by ERIC Clearinghouse, [Washington, D.C.], 1976.

Address for correspondence

Professor Karin Verspoor, karin.verspoor@unimelb.edu.au
The University of Melbourne, Melbourne VIC 3010 Australia.

Developing Methodologies to Find Abbreviated Laboratory Test Names in Narrative Clinical Documents by Generating High Quality Q-Grams

Kyungmo Kim^a, Jinwook Choi^{a, b}

^a Interdisciplinary program for Bioengineering, Seoul National University, Seoul 03080, South Korea,

^b Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul, 03080, South Korea

Abstract

Laboratory test names are used as basic information to diagnose diseases. However, this kind of medical information is usually written in a natural language. To find this information, lexicon based methods have been good solutions but they cannot find terms that do not have abbreviated expressions, such as “neuts” that means “neutrophils”. To address this issue, similar word matching can be used; however, it can be disadvantageous because of significant false positives. Moreover, processing time is longer as the size of terms is bigger. Therefore, we suggest a novel q-gram based algorithm, named modified triangular area filtering, to find abbreviated laboratory test terms in clinical documents, minimizing the possibility to impair the lexicons’ precision. In addition, we found the terms using the methodology with reasonable processing time. The results show that this method can achieve 92.54 precision, 87.72 recall, 90.06 f1-score in test sets when edit distance threshold(τ)=3.

Keywords:

Medical Informatics; Medical Informatics Computing; Natural Language Processing

Introduction

There are several ways to find similar words. The most representative case would be edit distance. However, its calculation cost is significant; therefore, existing studies apply q-gram based filters to filter the words that are not most likely to be involved the query term preferentially. Q-gram refers to substrings of words moving a window that is the size of q. For example, “folate” can generate q-grams of “fol”, “ola”, “lat”, “ate” where q=3.

Counter filtering can filter out terms that do not have common grams less than threshold of counter filtering= $(\max(|s|, |t|) - q + 1) - q \cdot \tau$. For example when q=3 and edit distance threshold(τ)=1, if “folder” generates “fol”, “ole”, “lde” and “der”, threshold of counter filtering would be 1. There is a common q-gram between terms “folate” and “folder”; hence, they are treated as similar terms. The terms that passed this test successfully were compared to the query word by using the edit distance calculation in next step.

However, counter filtering approach has an error. In its formula, “ $(\max(|s|, |t|) - q + 1)$ ” means the size of q-grams and “ $q \cdot \tau$ ” is the limit of differences between two words. Therefore, threshold of counter filtering means that the least common q-grams to judge two words are the same. If words are short, the threshold can become negative and generate a massive false positive. Therefore, in this study, we modify algorithms to generate less false positives and to make them faster.

Methods

Developing Algorithms

We describe counter filtering in 2-dimensional space. In Figure 1, the x-axis represents the number of q-grams and the y-axis represents common q-grams with the query term. The query word draws a virtual line, the “standard line” on both figures. If distances between the point of the target word and standard line reach a threshold line or over it, the target word is filtered out.

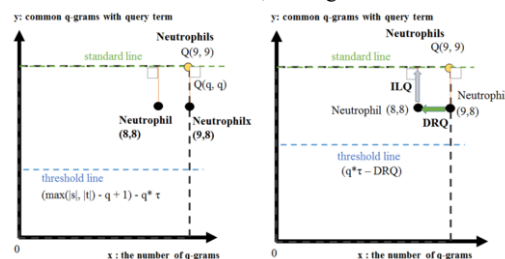


Figure 1 – Left: Convert q-gram based algorithm to 2-dimensional plane (traditional counter filtering); Right: Convert q-gram based algorithm to 2-dimensional plane (modified counter filtering).

Modified Q-Gram Filtering

This filtering concentrates on distances between two words from the perspective of q-grams. For example, consider q-grams of query word (QQW) : neutrophils = [neu, eut, utr, tro, rop, oph, phi, hil, ils] q-grams of target word (QTW) : neutrophilx = [neu, eut, utr, tro, rop, oph, phi, hil, ilx]. The different q-grams are “ils”, “ilx”. Here, $q \cdot \tau$ is the limit of differences between two words. Therefore, distances between two of them should be equal or less than $q \cdot \tau$. Thus, the condition to judge them as similar is $(QQW - CQ) + (QTW - CQ) \leq q \cdot \tau$. (CQ: common q-gram). We expressed equations $QQW - CQ$ as ILQ (insertion of lacked q-gram), $QTW - CQ$ as DRQ () for convenience. Here, to simplify the problem as vertical distance, the equation should be changed as follows: $ILQ \leq q \cdot \tau - DRQ$

Orthogonal Distance Filtering

A modified q-gram generates a massive false positive when it finds “neuts” as “neutrophils”. Therefore, orthogonal distance filtering was suggested to solve this problem. In this algorithm, substrings from the 0-th to i-th character in query words are treated as exemplary terms such as “neu”, “neut”, and “neutr”. Therefore, the standard line in Figure 1 tilts 45 degrees to become Figure 2. For this reason, all distances are multiplied by $1/\sqrt{2}$. This is shown in Figure 2 (left).

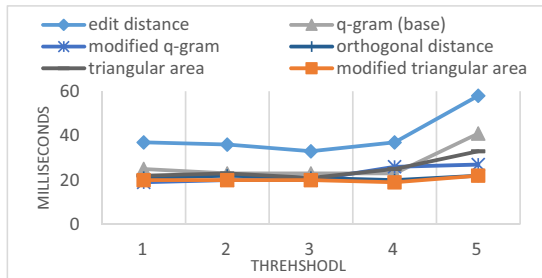


Figure 4 – Procedure time to proceed 2000 words.

Discussion

For a long time, there have been attempts to reduce the time of similar string matching. [1] reduced processing time by building a special inverted index. [2] concentrated on how to make a systemic inverted index; therefore, introducing a graph shaped inverted index. [3] concentrated on mismatching q-grams to speed up the algorithm. [4] split words, then applied prefixes to words to generate similar strings in documents. These works are not interested in precision but the increased speed of the algorithm, because the q-gram has been treated like as a preprocessing-step before edit distance. However, a massive false positive in candidate set 1 can raise the cost of the rest of the calculation steps like edit distance. Therefore, we developed our proposed algorithm.

Although the suggested algorithm used a more strict threshold, it could cultivate entities, many as existing algorithms with less false positives and within a reasonable time. This algorithm can be helpful to researchers who want to cultivate features from documents so that the burden on subsequent processing such as the validation step and semantical classification can be reduced, as the number of input entities was reduced.

Conclusion

We developed a novel algorithm successfully by modifying existing filtering algorithms based on q-gram techniques. The modified triangular area filtering and triangular area filtering generated good quality q-grams even cultivating abbreviated laboratory terms within moderate time. These algorithms have substituted existing counter filtering and edit distance because they are faster and generate better quality words. This work will be helpful to researchers who want to find similar strings in documents rapidly with lower calculation time and fewer candidate words.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIP) (No. 2010-0028631) and was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01058075). Deidentified clinical records used in this research were provided by the i2b2 National Center for Biomedical Computing funded by U54LM008748 and were originally prepared for the Shared Tasks for Challenges in NLP for Clinical Data organized by Dr. Ozlem Uzuner, i2b2 and SUNY.

References

- [1] D. Bahle, H. E. Williams, and J. Zobel, Efficient Phrase Querying with an Auxiliary Index, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 2002, 215-221.
- [2] K.M. Hammouda and M. S. Kamel, Phrase-based document similarity based on an index graph model, *Proceedings of the IEEE 2002 international conference on data mining (ICDM'02)*, 2002, 203-210.
- [3] C. Xiao, W. Wang, and X. Lin, Ed-Join: an efficient algorithm for similarity joins with edit distance constraints, *Proceedings of the VLDB Endowment*, 2008, 933-944.
- [4] W. Wang, C. Xiao, X. Lin, and C. Zhang, Efficient Approximate Entity Extraction with Edit Distance Constraints, *Proceeding SIGMOD '09 Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, 2009, 759-770.

Address for correspondence

Jinwook Choi, M.D., Ph.D.
 Department of Biomedical Engineering
 Seoul National University
 718 Basic Science Building
 Yongon Campus 28 Yongon-dong
 Chongro-gu
 Seoul 110-799
 Republic of Korea
 E-mail : jinchoi@snu.ac.kr

Appendix of supplementary information

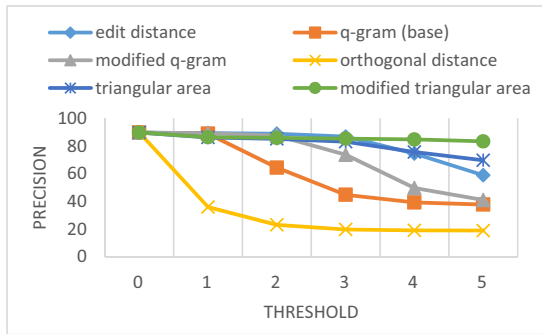


Figure 5–Precision trends in i2b2 2014 training set

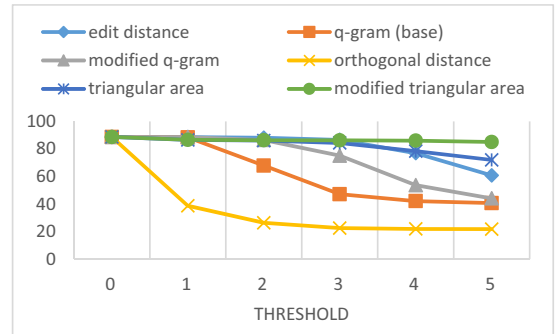


Figure 9–Precision trends in i2b2 2014 test set

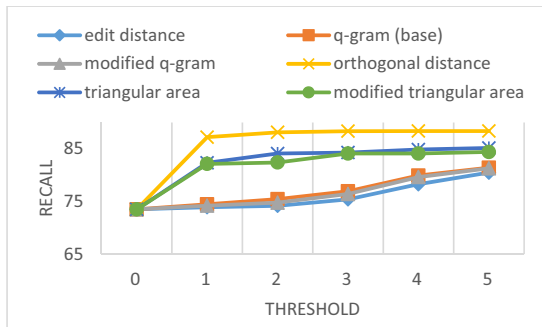


Figure 6–Recall trends in i2b2 2014 training set

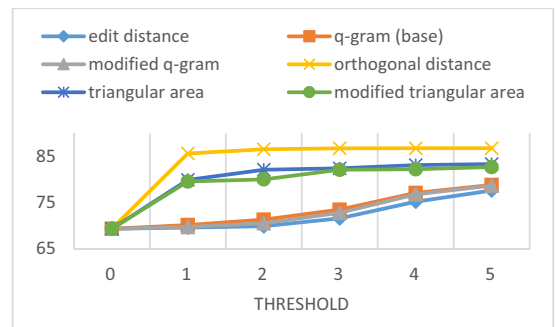


Figure 10–Recall trends in i2b2 2014 test set

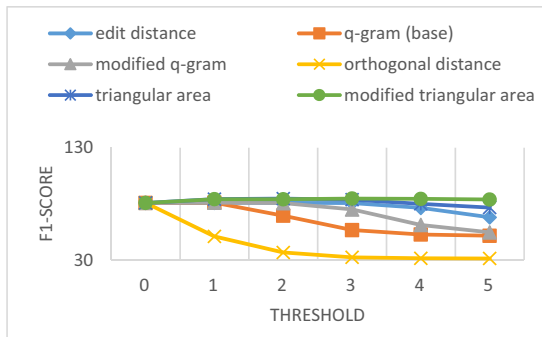


Figure 7–f1-score trends in i2b2 2014 training set

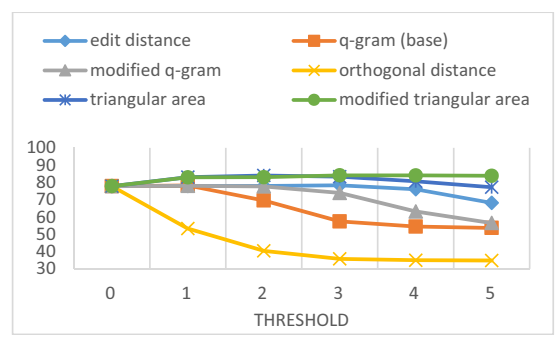


Figure 11–f1-score trends in i2b2 2014 test set

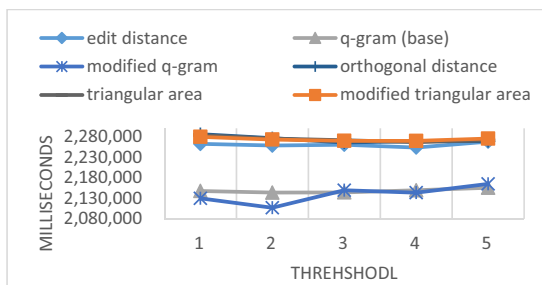


Figure 8 – procedure time in i2b2 2014 training set

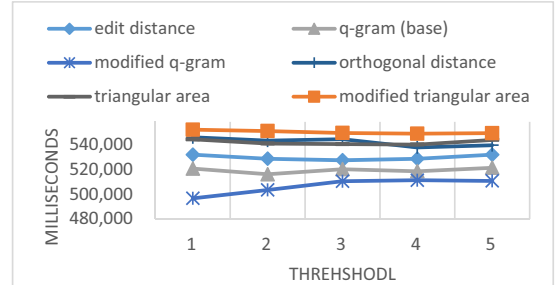


Figure 12–f1-score trends in i2b2 2014 test set

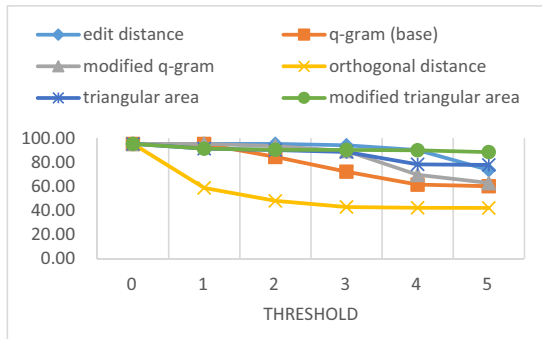


Figure 13–Precision trends in smuh_diabetes training set

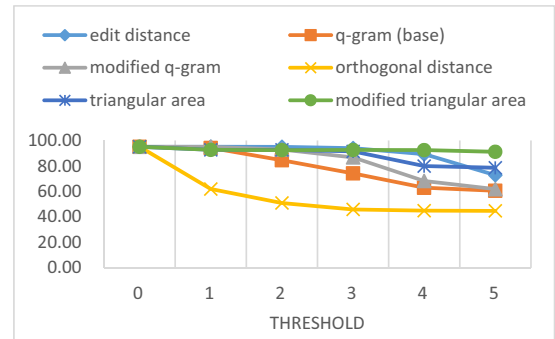


Figure 17–Precision trends in smuh_diabetes test set

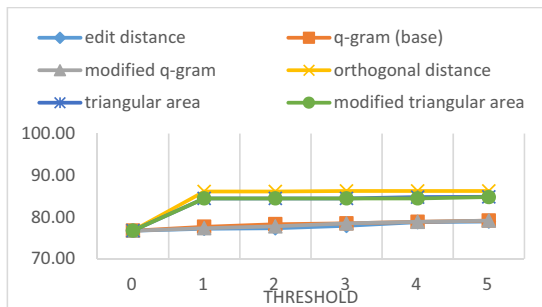


Figure 14–Recall trends in smuh_diabetes training set

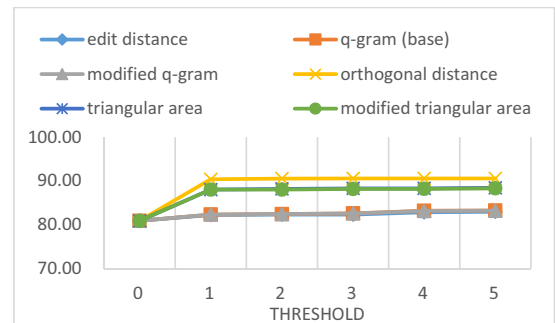


Figure 18–Recall trends in smuh_diabetes test set

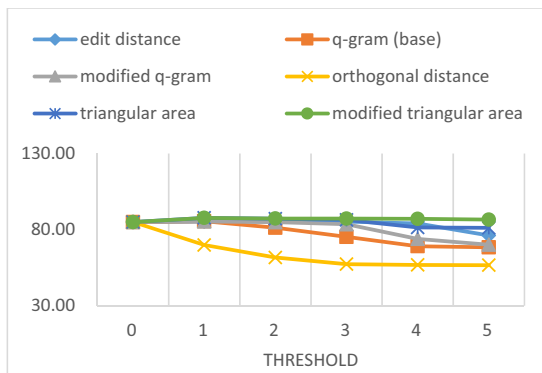


Figure 15–f1-score trends in smuh_diabetes training set

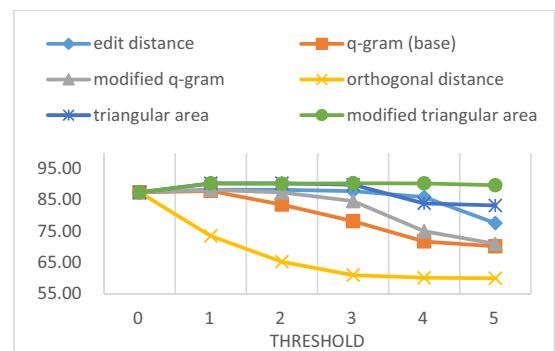


Figure 19–f1-score trends in smuh_diabetes test set

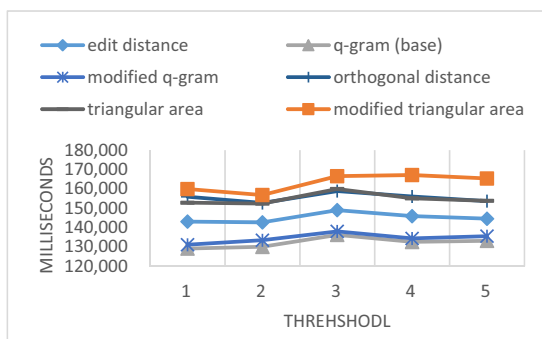


Figure 16 – procedure time in smuh_diabetes training set

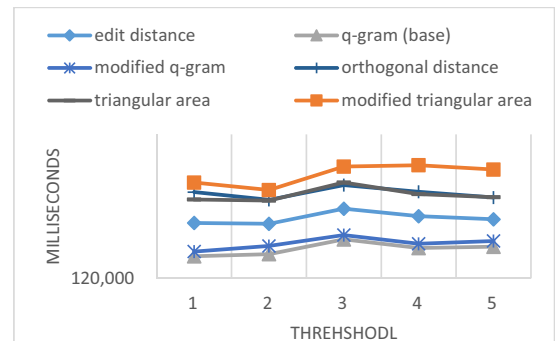


Figure 20–f1-score trends in smuh_diabetes test set

Characterising the Scope of Exposome Research: A Generalisable Approach

Philip Kiossoglou^a, Ann Borda^a, Kathleen Gray^{a,c}, Fernando Martin-Sanchez^{a,b},
Karin Verspoor^{a,c}, Guillermo Lopez-Campos^{a,d}

^a Health and Biomedical Informatics Centre, The University of Melbourne, Parkville, Victoria, Australia,

^b Environmental and Participatory Health Informatics (ENaPHI) Research Group, Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, NY, USA,

^c School of Computing and Information Systems, The University of Melbourne, Parkville, Victoria, Australia,

^d Centre for Experimental Medicine, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom

Abstract

Scientific advancement and the development of new research fields bring uncertainties about what the current topics of research emphasis are and thus, what new knowledge might need to be represented. The exposome is an example of one such new field for which these uncertainties exist. The exposome is the analogue to the genome, from an environmental exposure perspective; research on the exposome has gained momentum only since 2011. In this work, we propose a generally applicable methodology that aims to characterise the landscape of a new research area based on linguistic analysis of its associated publications. Using abstracts of 261 exposome research articles, we illustrate a methodology that combines (1) inductive analysis based on word frequency counts, and term analysis to identify the topics, methods and applications of the new field and (2) deductive analysis using the NCBO Ontology Recommender to identify to what extent this new area is covered by current knowledge representation tools. Applying this method to the exposome literature, we uncover both the current focus of exposome research and the ontologies that are most relevant to the domain.

Keywords:

Environmental Exposure; Medical Informatics, Text Mining, Biomedical Ontologies

Introduction

The role of environmental factors in health and disease along with the controversial “nature versus nurture” debate have been ongoing themes in science and research for over a century [1], [2], [3]. With the development of precision medicine, the relevance of both themes and the need to consider both in research have been explicitly acknowledged [4], [5]. This represents an attempt to satisfy the equation “Phenotype = Genome x Environment”. The “nature” component of this equation, the genome, has experienced a formidable boost in the last two decades. It has also fostered the development of what are known as “omics” approaches (proteomics, metabolomics, etc.) to characterize molecular phenotypes. A commonality among these approaches has been their reliance on recent technological advances to generate very large data sets. On the other hand, the “environmental” aspect of the equation was typically studied with population or environmental health approaches, until 2005 when C.P. Wild coined the term “Exposome” and further defined it as an exposure-oriented analogue of the genome [6]. The exposome represents the sum total of exposures an individual receives

over time, from both internal and external sources. Similarly to what has happened with the development of other “omics” approaches, study of the exposome is benefitting from recent advances in technology, especially with the development and reduction in cost of sensors. These new technologies have opened the door to the development of new strategies that shift research focus from population-based exposure assessment to more individualised approaches [7].

The concept of the exposome was initially established over a decade ago, and has gained significant attention only recently with the establishment of several large research projects [8-11], due in part to its demonstrated relevance to the new paradigm of “precision medicine” [12]. As a consequence of this increased focus, the body of knowledge and the volume of literature associated with the exposome have grown continuously over the last six years (Figure 1). Although still young, this is a rapidly evolving research field.

PubMed “Exposome” articles (per year)

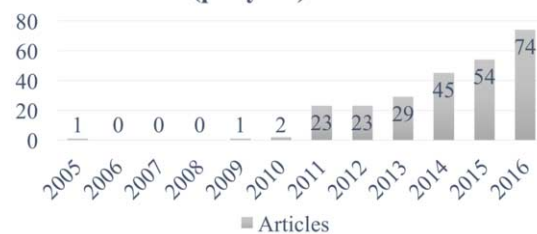


Figure 1 – Number of Articles Indexed in PubMed Containing the Term “exposome” since it was Coined in 2005.

The exposome is posing new challenges for biomedical informatics and requires the adaptation and development of methods and tools [13]. Knowledge representation methods, such as ontologies, have greatly supported “omics” research; examples of successful informatics tools include the Gene Ontology [14] and the Human Phenotype Ontology [15]. These ontologies provide formal vocabularies to describe knowledge and represent the relationships among entities within those vocabularies.

Exposome research is a relatively new field with an increasing corpus of literature but it is not yet well characterised in terms of formal knowledge representation, hampering our ability to develop suitable informatics tools. For this reason, in this work we propose a methodology to characterise this new research landscape, by identifying the terms and concepts that better

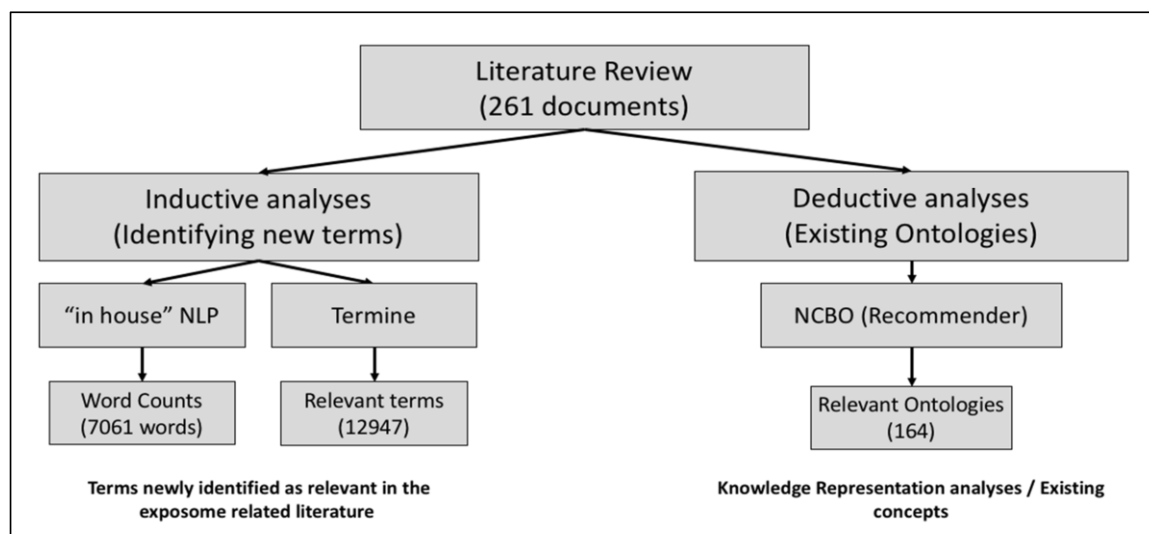


Figure 2 – Methodology for Characterising Exposome Research through Inductive and Deductive Analysis of the Literature.

describe exposome related knowledge. We believe that this methodology is applicable not only to the exposome but also is transferable to any new or developing research field.

Methods

Our characterization of the exposome research landscape is based on a literature review to discover the extent of relevant articles, followed by two different and complementary approaches to content analysis of their abstracts (Figure 2):

- An inductive approach based on text mining to identify the most relevant words and terms;
- A deductive approach based on identifying the most relevant ontologies, i.e. those that offer the best coverage of terms.

The literature selection was performed by searching for papers published between January 2005 and April 2016 with the term "exposom*" in the title, abstract or keyword fields. Searches were undertaken in the CABI, CINAHL, EMBASE, PubMed and Web of Science databases. Articles were excluded if they were written in a language other than English, if they were an e-book, or if they were related to research in non-human subjects. 22 items were excluded under these criteria, resulting in a total of 261 documents that were finally used in these analyses. Abstracts of the selected documents were then retrieved and stored in a database to form a corpus for analysis.

The inductive stage of the analysis was performed applying two different strategies. In the first strategy, "technical terms" were identified in the corpus using Termine [16], a tool that approaches novel multi-word term identification using statistical methods. The second strategy consisted of calculating word frequency counts in the corpus applying a natural language processing (NLP) protocol developed "in house" using Python's Natural Language Tool Kit (NLTK) [17].

Using Termine, abstracts were concatenated into a single file in order to provide enough text for effective processing in one pass. Prior to this concatenation, a preprocessing step on the

individual abstracts was carried out to improve the final outcomes and the submission process, as follows: Copyright notices were removed from the text; they were converted to lowercase and they were lemmatized to reduce the number of duplicated terms (for example "exposure" and "exposures" were grouped together); and finally sentences were tokenized (using the NLTK method) and replaced with new lines.

Word frequency analysis was performed in the following fashion. Abstracts were concatenated first, then converted to lowercase, and punctuation characters were removed. A frequency distribution was then produced using the `nlk.probability.FreqDist` procedure. This result was stored and then manually curated to limit the terms considered potentially relevant (for example, discarding closed-class function words such as 'the', 'of' and 'by').

For the deductive analysis, we used the repository of biomedical ontologies found at the National Center for Biomedical Ontology (NCBO). This resource currently catalogues more than 500 ontologies. The NCBO Ontology Recommender tool API [18] was applied in order to identify what ontologies covered this new field. An automatic process was used to pass each abstract separately to the API and record the ontologies recommended in each response.

Results

Inductive Analyses

Word count analysis identified 7061 unique words with a range of counts between 1 and 2844. The count distribution shows a Zipfian-like distribution, where only 12.7% (897) of the words have a word count greater or equal than 10. We further filtered these results applying a minimal word-count threshold of 10 followed by manual filtering of the remaining words. This filtering step, designed to remove frequent general English words and other words irrelevant for this field based on our own linguistic knowledge, resulted in a final list of 427 relevant words with word frequencies between 10 and 614. Of these 427 words, the top 25 showed a word frequency >100 (Table 1).

Table 1 – Top 25 Words in Exposome Abstracts (word counts >100) identified using word frequency analysis.

Word	Count	Word	Count
Exposure	614	Biomarker	151
Environmental	405	Analysis	136
Study	386	Effect	135
Disease	339	Cancer	133
Exposome	289	Method	128
Health	252	Genetic	118
Human	239	Molecular	115
Risk	215	Biological	114
Data	200	Individual	112
Factor	190	Development	104
Research	172	Assessment	101
Approach	166	Interaction	101
Chemical	165		

Termine identifies complex and technical terms comprising two or more words. These results are then ranked according to their frequency and an internally calculated score (c-value) that is derived from statistical and linguistic information [16]. This analysis resulted in the identification of 12,947 terms with c-values ranging between -1 and 105.8 and frequencies between 1 and 107. Of those terms identified, 95% occurred at a very low frequency (≤ 2), and only 1% had a frequency of ten or greater. Table 2 shows the top 25 terms identified.

Table 2 – Top 25 Multi-Word Terms in Exposome Abstracts, as Identified by Termine.

Term	c-value	Frequency
environmental exposure	105.8	107
environmental factor	50.9	52
risk factor	43.4	45
risk assessment	37.8	39
chronic disease	33.9	35
public health	33.9	35
association study	30.3	32
exposure science	28	29
gene expression	26.9	28
exposure assessment	23.9	25
environmental health	22.7	24
gene environment interaction	20.7	22
mass spectrometry	19.9	21
epidemiological study	19.8	21
disease risk	19.8	21
genome wide association study	19.0	13
human health	18.9	20
metabolic profiling	17.8	19
human exposome	17	18
air pollution	15.9	17
complex disease	15	16
human genome	14.8	16
human disease	14.6	16
omic technology	14	15
cohort study	13.7	15

Deductive Analysis

The NCBO Ontology Recommender tool uses four separate scores for each ontology based on internal calculations measuring the “coverage”, “detail of knowledge”, “specialization” and “acceptance” of the proposed ontology. These four categories are weighted to produce a default ranking. Using the NCBO default settings, the weights

employed are Coverage: 0.55, Detail of Knowledge: 0.15, Specialization: 0.15 and Acceptance: 0.15. These analyses returned a list limited to the twenty five (25) most highly ranked ontologies for a supplied abstract.

A total of 164 different ontologies were recommended, across the 261 abstracts. To better interpret these results we calculated a measure, the ontology recommendation frequency percentage (ORFP), to represent how often an ontology was recommended for this dataset. Applying an arbitrary threshold of ORFP > 50, only 17 (~10%) of the 164 suggested ontologies were recommended for more than 50% of abstracts (Figure 3 and Table 3).

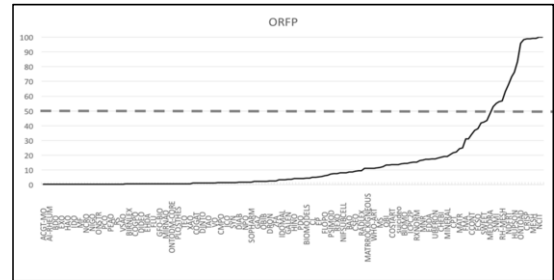


Figure 3 – Recommendation Frequency for all 164 Ontologies Recommended by the NCBO Ontology Recommender. The Dotted Line Shows the 50% ontology Recommendation Frequency Percentage (ORFP) Threshold. Not all Ontologies are Named on the X-Axis.

To further refine these results, we aggregated all 164 recommended ontologies and calculated a new combined ranking for the 17 ontologies above the threshold. This was performed by means of computing a “corrected rank” for each ontology, taken as the average rank it received for each individual abstract. If an ontology was not recommended for a given abstract, it was given an arbitrary rank of 26.

Table 3 – 17 ontologies recommended by the NCBO Ontology Recommender with ORFP > 50, showing their average and corrected ranks.

Ontology	ORFP	Average rank	Corrected rank
NCIT	100.00	1.16	1.16
SNOMEDCT	100.00	3.62	3.62
MESH	99.23	3.20	3.38
NIFSTD	99.23	7.37	7.51
CRISP	98.85	5.89	6.12
RCD	98.85	6.25	6.48
LOINC	98.47	5.63	5.94
ONTOAD	95.79	9.32	10.03
EFO	83.52	12.73	14.92
HUPSON	76.25	14.53	17.25
HL7	73.18	12.64	16.22
NDFRT	67.82	14.67	18.31
EDAM	63.22	14.85	18.95
RH-MESH	56.70	12.46	18.32
GO	56.32	14.37	19.45
SNMI	55.17	15.60	20.26
MEDLINEPLUS	52.87	13.51	19.40

To check for consistency between our inductive and deductive results, we added a deductive step where we used the list of multi-word terms filtered from Termine as input to the NCBO Ontology Recommender tool. This resulted in 25 recommended ontologies, all of which also appeared in the list of 164

ontologies derived from the deductive analysis of complete abstracts. These 25 ontologies matched 15 out of the 17 ontologies with an ORFP >50, demonstrating the overall representativeness of the inductive terms as compared to the broader literature. The two missing ontologies were EDAM and RH-MESH (Figure 4).

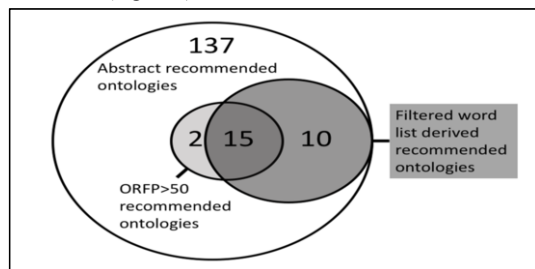


Figure 4 – Consistency Between Two Sets of Ontologies Recommended for Exposome Abstracts: 25 Recommended for Filtered Multi-Word List from Inductive Analysis, and 17 Recommended from Deductive Analysis of Whole Abstracts.

Discussion

The proposed methodology is based on two different and complementary approaches, an inductive strategy aiming to unveil unknown aspects associated with the terms and words (eventually concepts) that might characterise exposome research, and a deductive method that attempts to identify what existing elements in the biomedical knowledge representation space can be used to annotate exposome related published research.

The two inductive analyses were complementary to each other. The word frequency method resulted in the identification of individual words such as “environmental”, “exposure” or “biomarker” whereas the use of Termine focused on the identification of more complex terms and concepts such as “environmental health”, “environmental exposure” or “metabolic profiling”.

The word frequency approach was applied to identify the most frequent words in the literature and gave insights into individual elements that provide an overall perspective on the field. Although Table 1 shows only the top 25 results, this analysis enabled the identification of other relevant words in this context that show different areas where research has focused such as “epigenetic”, “metabolomic”, “maternal” or “lead” (a metal) and even more specialised words such as “adductomics”.

The second inductive approach, using Termine, allowed us to identify more complex terms and concepts that elude the simple word count approach. The reason why Termine analysis identified a large number of terms, greater than the individually identified words, is because the results include terms of different length (word *n*-grams, or multi-word terms) that are sequences of individual words (i.e. “environmental factor”, “environmental factor environome” and “environmental factor such investigation”). The results in Table 2 clearly show the complementarity to those identified by word count. The analysis of terms allowed us to identify methodologies (such as “cohort studies” or “genome wide association studies”) and techniques (“mass spectrometry” or “metabolic profiling”) that are relevant in exposome research. They thus identified terms which correspond to important concepts in this emerging field.

The approach based on deductive analysis using the ontology recommender allowed us to contextualise current formal knowledge representation of the exposome. Based on this analysis, only a small portion of the overall recommended

ontologies are broadly useful in this field of research at this point. More importantly, none of the ontologies that were frequently recommended had major associations with the science of environment or exposures, suggesting that there is a need to bridge interdisciplinary research gaps to build knowledge in this field.

Limitations

In spite of having successfully applied this methodology to the characterisation of exposome research with complementary and consistent results, this study has several limitations in its current form. One of these is connected with the use of abstracts rather than full text content in the analysis, since abstracts highlight the key aspects of the full-text content of the document but are not intended to be exhaustive summaries. However, since many full text publications are not available for analysis, this limitation is balanced with more exhaustive coverage of relevant literature using abstracts.

Another limitation lies in the use and interpretation of the results from the proposed automatic analyses with regard to the ambiguity of terms; the tools do not have the capacity to disambiguate different meanings. For example, in the inductive analysis the word “lead” would be equally counted regardless of whether it is acting as a noun (a metal) or as a verb (to precede). In the deductive analysis there is a similar problem with different ontologies recognising the same term but without contextualising it. In the aforementioned example “lead” is defined by NCIT as “Be in charge of; a position of leadership” whereas in CRISP it is defined as a “Soft grayish blue metal with poisonous salts; symbol, Pb, atomic number 82”.

General Application of the Approach

Scientific advancement and the development of new research areas bring uncertainties about how new knowledge can be or should be represented. The methodology proposed here enables objective assessment of such questions and thus facilitates analysis of the evolution of new fields of biomedical science, including those driven by new technologies.

For example in this study of exposome research, the deductive analysis showed the relevance of toxicogenomics in exposome research in the form of a number of genomic and bioinformatics related ontologies (GO, EDAM, EFO and HUPSON).

Our suggested method combining induction and deduction is generalisable and suitable to be applied in corpora of literature derived from other biomedical fields. The inductive stage is “blind” to the origin of the corpus used, and independent of any existing knowledge resources. Therefore it can equally applied in any area with an identifiable body of literature, allowing the most relevant words and terms that characterise the new field to surface and thus providing insight into the current methods, applications and general interests in the field – as well as the gaps. On the other hand, the deductive stage relates the corpus to existing knowledge encoded in the form of ontologies, and maps the places where aspects of current knowledge overlap with research in the new field of interest, or where there is no intersection.

Future Work

In future work, we plan to analyse full texts of articles from the literature instead of just abstracts. In addition, more complex natural language processing strategies will be applied in order to reduce the disambiguation problem. These results will eventually be combined with the results generated from the deductive approach to improve the interpretation of the latter.

This will support eventual comparison between the results from the analysis of exposome research literature and the results of

analysis of representative bodies of literature from related fields (for instance, other omics research). This may determine which features of exposome research are particularly novel and thus may help to direct further research.

Conclusion

We have developed a methodology that combines inductive and deductive approaches for the characterisation of the landscape of terms or concepts and availability of formal knowledge representation tools in a relatively new area of biomedical research. When applying this methodology to the study of exposome research, results from the inductive and deductive approaches are consistent and complementary, allowing us to identify different sets of terms, concepts and ontologies that describe the current status of human exposome-related knowledge. This is a major first step towards enabling development of informatics tools to support systematic comprehensive integration of exposome data into precision medicine.

The use of automated tools and methods like the ones described in this work adds to the growing body of work to address the limitations of the current processes in systematic reviews of this kind [19; 20]. Our methods could enable the generation of an overall perspective of research scope, and characterisation of the most relevant topics under investigation, in new disciplines. The approach further facilitates the assessment of whether new formal knowledge representation tools are required for the new discipline.

References

- [1] F. Galton, *Hereditary genius : an inquiry into its laws and consequences*, Macmillan and co., London, 1869.
- [2] B.S. Burks, Foster Parent-Foster Child Comparisons as Evidence upon the Nature-Nurture Problem, *Proc Natl Acad Sci U S A* **13** (1927), 846-848.
- [3] B.T. Eiduson, S. Eiduson, and E. Geller, Biochemistry, genetics, and the nature-nurture problem, *Am J Psychiatry* **119** (1962), 342-350.
- [4] F.J. Martin-Sanchez and G.H. Lopez-Campos, The New Role of Biomedical Informatics in the Age of Digital Medicine, *Methods Inf Med* **55** (2016), 392-402.
- [5] D.P. Jones, Y. Park, and T.R. Ziegler, Nutritional metabolomics: progress in addressing complexity in diet and health, *Annu Rev Nutr* **32** (2012), 183-202.
- [6] C.P. Wild, Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology, *Cancer Epidemiol Biomarkers Prev* **14** (2005), 1847-1850.
- [7] M.J. Nieuwenhuijsen, D. Donaire-Gonzalez, M. Foraster, D. Martinez, and A. Cisneros, Using personal sensors to assess the exposome and acute health effects, *Int J Environ Res Public Health* **11** (2014), 7805-7819.
- [8] C. Potera, The HELIX Project: tracking the exposome in real time, *Environ Health Perspect* **122** (2014), A169.
- [9] M. Vrijheid, R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T.J. Athersuch, N. Avellana, X. Basagana, C. Brochet, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J.R. Gonzalez, B. Granum, R. Grazuleviciene, K.B. Gutzkow, J. Julvez, H.C. Keun, M. Kogevinas, R.R. McEachan, H.M. Meltzer, E. Sabido, P.E. Schwarze, V. Siroux, J. Sunyer, E.J. Want, F. Zeman, and M.J. Nieuwenhuijsen, The human early-life exposome (HELIX): project rationale and design, *Environ Health Perspect* **122** (2014), 535-544.
- [10] HERCULES Exposome Research Center, <http://emoryhercules.com/>
- [11] P. Vineis, M. Chadeau-Hyam, H. Gmuender, J. Gulliver, Z. Herceg, J. Kleinjans, M. Kogevinas, S. Kyrtopoulos, M. Nieuwenhuijsen, D.H. Phillips, N. Probst-Hensch, A. Scalbert, R. Vermeulen, C.P. Wild; EXPOsOMICS Consortium, The exposome in practice: Design of the EXPOsOMICS project, *Int J Hyg Environ Health* **16** (2016), 30130-4.
- [12] National Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease., *Toward precision medicine : building a knowledge network for biomedical research and a new taxonomy of disease*, National Academies Press, Washington, D.C., 2011.
- [13] F. Martin Sanchez, K. Gray, R. Bellazzi, and G. Lopez-Campos, Exposome informatics: considerations for the design of future biomedical research information systems, *J Am Med Inform Assoc* **21** (2014), 386-390.
- [14] C. The Gene Ontology, Expansion of the Gene Ontology knowledgebase and resources, *Nucleic Acids Res* (2016).
- [15] S. Kohler, N.A. Vasilevsky, M. Engelstad, E. Foster, J. McMurphy, S. Ayme, G. Baynam, S.M. Bello, C.F. Boerkoel, K.M. Boycott, M. Brudno, O.J. Buske, P.F. Chinnery, V. Cipriani, L.E. Connell, H.J. Dawkins, L.E. DeMare, A.D. Devereau, B.B. de Vries, H.V. Firth, K. Freson, D. Greene, A. Hamosh, I. Helbig, C. Hum, J.A. Jahn, R. James, R. Krause, S.J. Laulederkind, H. Lochmuller, G.J. Lyon, S. Ogishima, A. Olry, W.H. Ouwehand, N. Pontikos, A. Rath, F. Schaefer, R.H. Scott, M. Segal, P.I. Sergouniotis, R. Sever, C.L. Smith, V. Straub, R. Thompson, C. Turner, E. Turro, M.W. Veltman, T. Vulliamy, J. Yu, J. von Ziegenweidt, A. Zankl, S. Zuchner, T. Zemojtel, J.O. Jacobsen, T. Groza, D. Smedley, C.J. Mungall, M. Haendel, and P.N. Robinson, The Human Phenotype Ontology in 2017, *Nucleic Acids Res* (2016).
- [16] K. Frantzi, Ananiadou, S. and Mima, H., Automatic recognition of multi-word terms: the C-value/NC-value method, *International Journal on Digital Libraries* **3** (2000), 15.
- [17] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*, O'Reilly, Beijing ; Cambridge Mass., 2009.
- [18] C. Jonquet, M.A. Musen, and N.H. Shah, Building a biomedical ontology recommender web service, *J Biomed Semantics* **1 Suppl 1** (2010), S1.
- [19] A. O'Mara-Eves, J. Thomas, J. McNaught, M. Miwa, and S. Ananiadou, Using text mining for study identification in systematic reviews: a systematic review of current approaches, *Syst Rev* **4** (2015), 5.
- [20] M. Verdonck, F. Gailly, S. de Cesare, and G. Poels, Ontology-driven conceptual modeling: A systematic literature mapping and review, *Applied Ontology* **10** (2015), 197-227.

Address for correspondence

Guillermo Lopez-Campos

Email : guillermo.lopez@unimelb.edu.au

Creation of the First French Database in Primary Care Using the ICPC2: Feasibility Study

Lacroix-Hugues V^{a,b}, Darmon D^{a,e}, Pradier C^b, Staccini P^{c,d,e}

^a *Deptt d'Enseignement et de Recherche en Médecine Générale, Université Nice Sophia-Antipolis, Nice, France*

^b *Département de Santé Publique, CHU de Nice, Nice, France*

^c *Département d'Information et d'Informatique Médicale, CHU de Nice, Nice, France*

^d *Département d'Ingénierie du Risque et Informatique de Santé, CHU de Nice, Nice, France*

^e *UMR 912 SESSTIM INSERM, Marseille, France*

Abstract

The objective of our study was to assess the feasibility of gathering data stored in primary care Electronic Health records (EHRs) in order to create a research database (PRIMEGE PACA project). The software for EHR models of two office and patient data management systems were analyzed; anonymized data was extracted and imported into a MySQL database. An ETL procedure to code text in ICPC2 codes was implemented. Eleven general practitioners (GPs) were enrolled as "data producers" and data were extracted from 2012 to 2015. In this paper, we explain the ways to make this process feasible as well as illustrate its utility for estimating epidemiological indicators and professional practice assessments. Other software is currently being analyzed for integration and expansion of this panel of GPs. This experimentation is recognized as a robust framework and is considered to be the technical foundation of the first regional observatory of primary care data.

Keywords

Electronic Health Records; Database; Information Dissemination

Introduction

Computerization of general practitioners (GPs) has been developed over the last 20 years [1]. The use of electronic health records (EHRs) and e-prescribing are priorities in different European action plans on e-health and are reflected in the policies of several states [2]. EHR development constitutes a formidable source of information and large data collection networks in primary care exist in our European neighbors, notably in the United Kingdom and the Netherlands (Clinical Practice Research Datalink CPRD Ex GPRD) [3], Q research [4], Netherlands Information Network in General Practice [5]). These development efforts have enabled the creation of voluminous databases allowing multiple publications.

In France, there is no similar model currently in operation. Consequently, unlike the hospital sector for which the established diagnoses are known, thanks to the PMSI (Program of Medicalization of the Information Systems), the morbidity supported in primary care is not well known [6].

The objective of our study was to assess the feasibility of gathering data stored in primary care EHRs in order to create a research database (PRIMEGE) PACA project: Plateforme

Régionale d'Information en Médecine Générale en Provence-Alpes-Côte d'Azur [7].

Methods

Data Definition

In a preliminary study, we identified the data provided by an EHR system. From the recommendations of the ANAES (Agence Nationale d'Accréditation et d'Evaluation en Santé) [8], concerning medical records in general medicine in addition to a study led by the departments of General Medicine at the University of Nice Sophia-Antipolis and Lyon 1 [9], we established a list of data to be collected. These data concern the episode of care, but also include data useful to evaluate and improve the practices (such as the ICPC2 code (international classification of primary care) enabling standardization of the reason for encounter and diagnoses). Data allowing patient identification (such as name, surname, address, etc.) were excluded.

Database Design

The modeling of a problem (i.e., the passage of the real world to its computerized representation), includes the integration of data into a relational database management system (SGBD-R) and allows for the manipulation of the data using SQL language (Structured Query Language) [10]. Classically, the data modeling process includes two phases: realization of a conceptual model and translation into a relational model. To establish the conceptual model, we used an entity-association (E-A) graphical formalism (widely used for the design of SGBD-R). We first built an E-A model centered on the visit (Figure 1). To be compatible with the the POMR (Problem Oriented Medical Record) software formalism, we built a second model that was centered on the problem of interest. The only difference between the first and second models was the visit table; start date and end date fields as well as a possible link to a parent problem were added. The two E-A models were then merged into the relational model (Figure 2). To link the CIP (Classification of Instructional Programs) code (drug table) with the corresponding ATC (Anatomical Therapeutic Chemical) class, a supplementary table (drug booklet) was added. The cross-reference table was supplied by Thériaque®. To be able to directly reach the reasons of encounter, diagnoses, and drugs prescribed for a given patient, additional access was added (represented by dashes on the relational model).

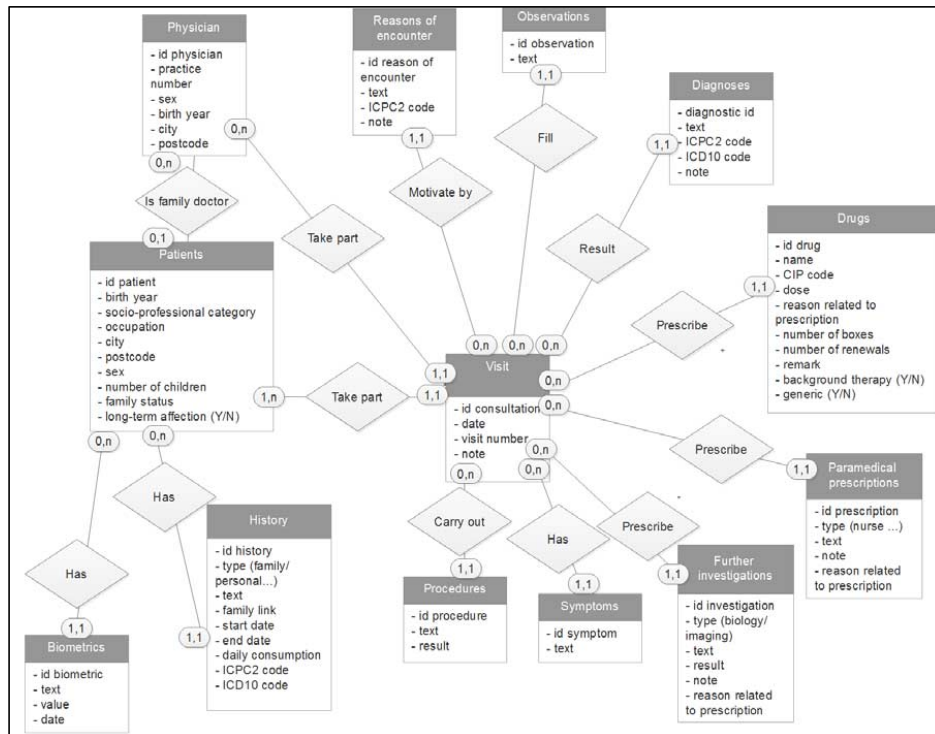


Figure 1 – Entity Association model centered on the visits

The PRIMEGE PACA database was developed under MySQL (version 5.1.73). PhpMyAdmin (version 3.5.2.2) was installed providing a graphical interface and to facilitate database management. The database is hosted at the CHU of Nice on a Linux SSH (Secure Shell) access. puTTY was installed to access the server in command mode and transferring files to the server is made via SFTP (SSH File Transfer Protocol).

Extraction Procedure

One of the most frequently used data management software systems by French GPs was analyzed using a workstation provided by a medical office [9]. Its structure was analyzed, and tables and fields of interest were identified. Once this work was completed, SQL queries to collect data were developed. To reduce interference with the GPs' work, extraction queries were run in batches, as background tasks, to generate needed .TXT and .CSV files.

Data extraction was realized using two methods. The first method involved the physician office where data were recovered on USB key. The second method was performed remotely using certified remote control software. This extraction procedure could be extended for use with other software.

Recruitment of the Investigators

Initially, internship supervisors working with one of the two software systems were contacted via the department of general medicine of the Faculty of Medicine of Nice. Upon agreement (all the partners of the contacted GPs agreed to participate), we went to their offices to explain the project in detail.

Data Importation

After quality checks performed, data were imported into intermediate tables corresponding to the structure of the software from which they were extracted. Once the data was imported into the intermediate tables, other scripts were used to convert the data into the PRIMEGE data model. During the process, copy tables were created, allowing the ability to return to the previous state in case any problems were encountered. Additionally, log files were generated to monitor the progress of the scripts and to analyze any errors encountered. A final script allowed copy tables to be removed and intermediate tables to be reset.

To ensure unique identifiers, a code corresponding to each medical office was concatenated onto the EHRs identifiers.

Automatic Coding

During extraction, few codes capturing the encounter and diagnoses reasons were found. "Coded" data are needed to standardize the ways that clinical concepts are represented. To solve this problem, we developed an automatic coding procedure based on the ICPC2 (recognized by 2003 as the international reference classification for primary care). This procedure compares the wording supplied with a list of strings (established from the reference document: treatment of medical information by the ICPC2 [11]) and stops at the first match found (Figure 3).

Legal Aspects

A declaration was made to CNIL (French supervisory authority for the protection of personal data registration no. 1585962). An informative poster intended for the patients and

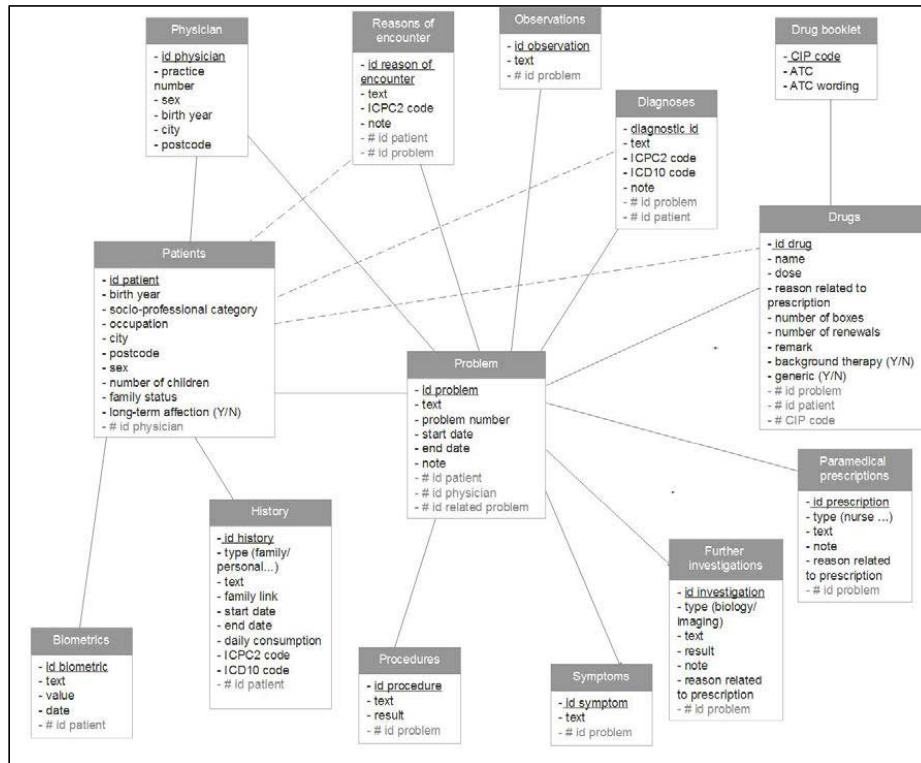


Figure 2 – Relational Model

```

ELSEIF v_libelle like ("%anemie%") then
  IF (v_libelle like ("%fer%") or v_libelle like ("%marial%")) then
    set v_CISP2 = 'B80';
    set v_CIM10 = 'D50';
  ELSEIF (v_libelle like ("%biermer%") or v_libelle like ("%b12%")) then
    set v_CISP2 = 'B81';
    set v_CIM10 = 'D51';
  ELSEIF (v_libelle like ("%folate%") or v_libelle like ("%b9%") or
    v_libelle like ("%foliq%")) then
    set v_CISP2 = 'B81';
    set v_CIM10 = 'D52';
  ELSEIF v_libelle like ("%grossess%") then
    set v_CISP2 = 'W84';
    set v_CIM10 = 'O99';
  ...
ELSE
  set v_CISP2 = 'B82';
  set v_CIM10 = 'D64';
END IF;
  
```

Figure 3 – Example of automatic coding

explaining the modalities of access and rectification of the data was arranged among the member physicians.

Results

Eleven physicians agreed to participate. The inclusions and data collection began in 2012. Finally, we collected four years of data (2012 to 2015) (Table 1, 2).

Table 1 – Data volume

Field	Number
Patients	33,940
Consultations	205,343
History	127,387
Biometrics	185,401
Reasons of encounter	111,151
Diagnoses	107,114
Drug prescription lines	485,947
Reported symptoms	432
Procedures	7,141
Further investigations	470,191
Paramedical prescriptions	7,964
Observations	24,527

Although reasons of encounter and diagnoses were poorly coded by GPs (14,60% and 1,88%), the automatic coding procedure allowed us to associate main reasons of encounter and diagnoses with an ICPC2 code (97,05% and 96,83%).

Table 2 – Patients characteristics (n = 33,940)

Characteristics	Number (%)
Age	
Childs (<18 years)	5,520 (16.26%)
Adults (18-64 years)	21,365 (62.95%)
Seniors (>64 years)	7,055 (20.79%)
Sex	
Women	17,973 (52.95%)
Men	15,456 (45.54%)
Not specified	511 (1.51%)

Discussion

Our study assessed the feasibility of gathering data stored in primary care EHRs in order to create a research database.

Strengths

Data Validity

To evaluate data quality, we considered the following criteria:

- Internal validity: 10 files per GP were randomly selected and searched for the corresponding data in the PRIMEGE database. In case of missing data, we returned to the extraction procedure to correct the problem.
- External validity: data volume was coherent since the number of annual consultations per GP was close to that indicated in the IRDES (Institut de Recherche et Documentation en Economie de la Santé) report on the GPs' working time (4667 vs 4225) [12].

Potential Impact

Secondary use of EHRs offer many perspectives:

- Clinical perspectives: These data could be used to improve quality of care, especially chronic disease management, [13] and could be used to assess professional practices. For the GPs, a better knowledge of the main problems encountered would allow for continuous learning. These data could also be used to evaluate the impact of the recommendations.
- Public health perspectives: Data from EHRs would facilitate public health surveillance by allowing for the estimation of epidemiological indicators and providing better data to guide public health interventions [14-15].

Interoperability

Our data model was conceived regardless of any medical software constraints. Intermediate table creation and data reprocessing to be compatible with the PRIMEGE model enables future adaptation to different medical software and ensures interoperability of the model.

Weaknesses

Automatic Coding Procedure Evaluation

Although the reliability of our automatic coding procedure was not formally established, an initial validation was carried out. For this, we compared the codes obtained with those found in ECOGEN (étude des Éléments de la Consultation en médecine GÉNérale) [16] as well as those provided by the CISMef (Catalogue et Index des Sites Médicaux de langue Française) team in Rouen using the same labels. In case of divergence between the codes, the reference document [11] was re-examined by analyzing the inclusion and exclusion criteria. In the future, an additional evaluation will be performed as part of a thesis (a representative sample of reasons of encounter and diagnoses will be manually coded and the results will be compared with those of the procedure to determine their accuracy).

Missing Data

The variability in completeness of data across patients and time requires careful consideration; restriction to those with

complete data may result in biased analyses [17]. One solution would be to encourage contributing practices to maintain a certain level of data quality through financial and/or intellectual incentives. In addition, some missing data could be collected via automatic cross-reference between prescriptions, laboratory results, and ICP2 codes [18]. Our automatic coding procedure could also be used to extract information from notes provided in free text fields by doctors.

Perspective

Other software is currently being analyzed for integration and expansion with this first panel of GPs. This regional model could be created as a network of regional observatories thus providing a representative panel of general practitioners' "data producers". Linkage to hospital data would enable the tracking of patient trajectories. To link PRIMEGE PACA data with hospital data we intend to use a probabilistic matching procedure as described in an IRDES' study [19].

Conclusion

This experimentation is recognized as a robust framework that could be adapted to increase data storage. This framework could be considered the technical foundation of the first French regional observatory of primary care data.

As far as we know, this is the first French data consolidation work in primary care. Currently, we use the ICP2, but other international classifications like SNOMED and ICD10 should complement this model to refine the results of epidemiological investigations and practice assessment.

References

- [1] S. De Rosi and C. Seghieri, Basic ICT adoption and use by general practitioners: an analysis of primary care systems in 31 European countries, *Bmc Medical Informatics and Decision Making* **15** (2015).
- [2] K.A. Stroetmann, J. Artmann, V.N. Stroetmann, D. Protti, J. Dumortier, S. Giest, U. Walossek, and D. Whitehouse, European countries on their journey towards national eHealth infrastructures - evidence on progress and recommendations for cooperative actions - Final European progress report, in, European Commission, DG Informations Society and Media, ICT for Health Unit, 2011.
- [3] CPRD (2017) The Clinical Practice Research Datalink (CPRD). Retrieved from <https://www.cprd.com>.
- [4] QReserach. Retrieved from <http://www.qresearch.org/>.
- [5] D.d. Bakker (2000) *Netherlands Information Network in General Practice (NIN GP)*. Retrieved from http://www.ulb.ac.be/esp/emd/nl_debakker.htm.
- [6] K. Chevreul, P. Le Fur, T. Renaud, and C. Sermet, *Faisabilité d'un système d'information public sur la médecine de ville.*, IRDES (Institut de recherche et documentation en économie de la santé), 2006.
- [7] PRIMEGE. Retrieved from <http://www.primege.org>.
- [8] La tenue du dossier médical en médecine générale. Etat des lieux et recommandations, in, Agence Nationale d'Accréditation et d'Evaluation en Santé, Paris, 1996.
- [9] D. Darmon, R. Sauvart, P. Staccini, and L. Letrilliart, Which functionalities are available in the electronic health record systems used by French general practitioners? An assessment study of 15 systems, *journal of medical informatics* **83** (2014), 37-46.
- [10] R. Chaabane, *Conception des bases de données: Modèle Entité-Association, Modèle Entité-Association*. Retrieved from <http://www.ai.univ-paris8.fr/~lysop/bd/seance2->

[ModeleEA.pdf](#).

- [11] M. Jamoulle, M. Roland, J. Humbert, and J. Brûlet, Traitement de l'information médicale par la Classification Internationale des Soins Primaires: CISP-2: deuxième version, *Bruxelles, Care Editions, ASBL* (2000).
- [12] P. Le Fur, Y. Bourgueil, and C. Cases, Le temps de travail des médecins généralistes: une synthèse des données disponibles, *Questions d'économie de la santé* (2009), 1-8.
- [13] S. de Lusignan and C. van Weel, The use of routinely collected computer data for research in primary care: opportunities and challenges, *Family Practice* **23** (2006), 253-263.
- [14] W.R. Hersh, Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance, *American Journal of Managed Care* **13** (2007), 277-278.
- [15] G.S. Birkhead, M. Klompas, and N.R. Shah, Uses of Electronic Health Records for Public Health Surveillance to Advance Public Health, *Annual Review of Public Health, Vol 36* **36** (2015), 345-359.
- [16] L. Letrilliart, I. Supper, M. Schuers, D. Darmon, P. Boulet, and M. Favre, ECOGEN: étude des éléments de la consultation en médecine général, *Exercer* **114** (2014), 148-157.
- [17] E. Herrett, A.M. Gallagher, K. Bhaskaran, H. Forbes, R. Mathur, T. van Staa, and L. Smeeth, Data Resource Profile: Clinical Practice Research Datalink (CPRD), *International Journal of Epidemiology* **44** (2015), 827-836.
- [18] C. Chmiel, H. Bhend, O. Senn, M. Zoller, T. Rosemann, and F.S. Group, The FIRE project A milestone for research in primary care in Switzerland, *Swiss Medical Weekly* **141** (2011).
- [19] Y. Bourgueil, J. Perlberg, C. Allonier, P. Boisenault, F. Daniel, P.L. Fur, and P. Szidon, Le rapprochement de données de médecine générale et de remboursement de l'Assurance maladie : étude de faisabilité et premiers résultats, *Questions d'économie de la santé, IRDES* **196** (2014), 0.

Address for correspondence

Virginie Lacroix-Hugues

Département d'Enseignement et de Recherche en Médecine Générale,
Université Nice Sophia Antipolis - Faculté de Médecine, 28 avenue
Valombrose 06107 NICE Cedex 2, France

Email : vhugues@outlook.

Conversion of National Health Insurance Service-National Sample Cohort (NHIS-NSC) Database into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM)

Seng Chan You^{a*}, Seongwon Lee^{b*}, Soo-Yeon Cho^a, Hojun Park^a,
Sungjae Jung^a, Jaehyeong Cho^a, Dukyong Yoon^a, Rae Woong Park^{a,c}

^aDepartment of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

^bCenter for Education Artificial Intelligence, Dankook University, Yongin, Korea

^cBrainKorea21 (BK21) Plus, Korea

*The first two authors contributed equally to this work

Abstract

It is increasingly necessary to generate medical evidence applicable to Asian people compared to those in Western countries. Observational Health Data Sciences and Informatics (OHDSI) is an international collaborative which aims to facilitate generating high-quality evidence via creating and applying open-source data analytic solutions to a large network of health databases across countries. We aimed to incorporate Korean nationwide cohort data into the OHDSI network by converting the national sample cohort into Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM). The data of 1.13 million subjects was converted to OMOP-CDM, resulting in average 99.1% conversion rate. The ACHILLES, open-source OMOP-CDM-based data profiling tool, was conducted on the converted database to visualize data-driven characterization and access the quality of data. The OMOP-CDM version of National Health Insurance Service-National Sample Cohort (NHIS-NSC) can be a valuable tool for multiple aspects of medical research by incorporation into the OHDSI research network.

Keywords:

Delivery of Health Care; Medical Informatics Application; Database

Introduction

Use of existing observation data to generate medical evidence is rapidly increased in terms of quantity of data, diversity of data sources, and the transparency [1]. Observational Health Data Sciences and Informatics (OHDSI) is an international collaborative consortium which aims to facilitate generating high-quality evidence via creating and applying open-source data analytic solutions to a large network of health databases across countries [2]. The OHDSI adopted distributed research network (DRN), which mediates observational studies to be conducted using multi-site database, while confidential personal health data remain within the original data holders [3]. The standardized same data structure, called the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM), is imperative to create network-wide results through the DRN by running the same analysis program for cooperating organizations. The OMOP-CDM was first developed in 2008 for drug surveillance study and has expanded its capacity to other research area. It supports

various types of studies such as drug/procedure safety, drug/procedure comparison, and medical cost analysis.

South Korea adopts a compulsory social insurance program, covering the whole population living in the country [4]. The National Health Insurance Service (NHIS), the institution for the Korean health insurance service holds the health claim database for all Koreans. In 2015, The NHIS released the NHIS-National Sample Cohort (NSC) database, which is a population-based sample cohort as the representative of the population as a whole.

In this paper, we present the converting process of the NHIS-NSC database into the OMOP-CDM and the result, focusing the executing phases.

Methods

Data Source

Korean public health insurance system for all citizens was initiated in 1963. Universal healthcare coverage was achieved in 1989. In 2000, the NHIS was launched as a single-insurer system by integrating more than 366 medical insurance organizations, for efficient system operation in Korea. The NHIS maintained national records for healthcare utilization and prescription over 98% of Korean whole population as of 2006. The NHIS established the NHIS-NSC, which was population-based cohort to provide representative, useful health insurance and health examination data to public health researchers and policy makers in 2015. About one million subjects, 2% of the Korean whole population, were selected by stratified random sampling from 2002 Korean health insurance database. Longitudinal health records in these population were collected for 11 years from 2002 to 2013. To preserve total number of subjects in the cohort, a representative sample of newborn was added annually as expired or emigrated subjects were excluded. The NHIS-NSC database can be assessed on the website [<http://nhiss.nhis.or.kr/bd/ab/bdaba021eng.do>].

Code Mapping

The NHIS uses Korean national medical code system. The 6th Korean standard Classification of Diseases (KCD-6) code is used for the diagnoses, which was originated from ICD-10 code. The NHIS has its own code system, electronic document

interchange (EDI) codes for drug, procedures and measurements.

The OMOP-CDM uses OMOP Standard Vocabulary(hereafter OMOP code), which requires transformation process from the local medial code system. OMOP code is based on RxNorm for medical drug and SNOMED-CT for medical diagnosis. Figure 1 shows how a drug code from Korean local drug code system is transformed to standard OMOP code. To consolidate the meaning of multi-site data, code transformation system is essential.

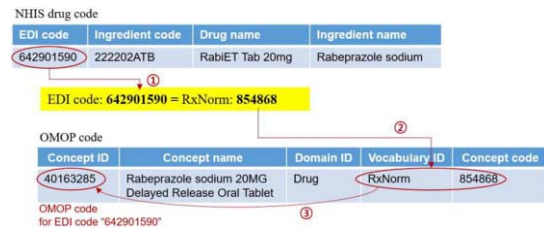


Figure 1– The example of transformation process from Korean local drug code to OMOP code.

We adopted the Korean Code Mapping Dictionary (KCMD) as a means of transformation from the NHIS codes to the OMOP codes. The KCMD is the code mapping dictionary between Korean local codes and the OMOP codes, which is developed by the authors. It shows 99% coverage for the diagnoses (KCD-6), 95% for drugs (EDI codes), and 99% for procedure (EDI codes) by the number of patients treated. For measurement, we additionally developed a mapping dictionary.

Extraction, Transformation, Loading (ETL) process

The OMOP-CDM embraces a variety of medical data such as patients, encounters, diagnoses, drug exposures, procedures, devices, results from laboratory tests, anthropometric measurements or questionnaire, and medical costs.

We converted the original data necessary for drug safety monitoring into OMOP-CDM with priority. Therefore, the seven tables in OMOP-CDM were generated: person, death, visit_occurrence, condition_occurrence, drug_exposure, procedure_occurrence, and location. With these converted tables, the various research including drug effect analysis, pharmacovigilance analysis can be performed.

The logical ETL mapping rules between the NHIS-NSC and OMOP-CDM were first defined. There were problems in developing ETL mapping documents. First, the NHIS-NSC provides subjects' age as a form of five-year-interval age group. Therefore, the approximate birth year of each subject had to be presumed. Second, all treatments such as drugs, procedures, and medical devices were stored in a single table in the NHIS-NSC and there was no reference data to accurately distinguish the type of treatments. We had to execute ETL by applying the KCMD for each treatment to source data.

We developed the physical ETL scripts as a form of Standard Query Language (SQL) and performed them as actual ETL process.

Data Characterization and Quality Management

ACHILLES (Automated Characterization of Health Information at Large-scale Longitudinal Evidence System) is a open-source OMOP-CDM-based data profiling tool, developed

by the OHDSI community. The ACHILLES creates a data-driven characterization and visualizes this result, by generating high-level aggregating statistics in a database based on the OMOP-CDM. The quality of the data can be assessed by the ACHILLES Heel, which is a subcomponent of the ACHILLES.

It is possible that mistakes in ETL processes occur [5]. To verify that the transformed data is equivalent to the original cohort data, we used statistical results from the ACHILLES to compare the population-level statistics with the previous reported cohort profile [6]. The ACHILLES Heel was also conducted to evaluation the data quality. The ACHILLES Heel generates two types of notifications: errors and warnings. "Errors" indicate more serious data quality issues, while "warnings" represent faults anticipated to have smaller impact.

Results

CDM Conversion Performance

Table 1 is the result of ETL, which shows counts of converted database records and conversion rates for OMOP-CDM tables. All data in the NHIS-NSC equivalent to subject, death, hospital visit and location were totally converted, resulting in 100% conversion rate. For the condition table, 89.4% diagnosis codes were mapped to SNOMED-CT and 99.9% of condition data were converted to OMOP-CDM with the mapping diagnosis codes. This means that 89.4% of codes showed the 99.9% prevalence in condition data. For drug and procedure, the total about 975 million data were converted to OMOP-CDM. Among them, drug and procedure comprised 53.6% (522,575,793 records) and 46.4% (452,147,182 records). Its whole conversion rate was 94.5%.

Table 1– Count of data converted and conversion rate from the NHIS-NSC to the OMOP-CDM

Table	Data count in data source	Data count converted	Conversion rate
Subject	1,125,691	1,125,691	100.0%
Death	55,921	55,921	100.0%
Hospital visit	119,362,188	119,362,188	100.0%
Condition	299,379,695	299,053,439	99.9%
Drug/Procedure	1,031,760,325	974,722,975	94.5%
Location	317	317	100%

Data Characteristics and Visualization

The ACHILLES estimated that total number of subject in the cohort was 1.13 million. Total of 560,640 (49.8%) subjects were female (Figure 2). The reported number of all subjects and annual included infants in the cohort are compared with the result from the ACHILLES (Table 2).

The number estimated by the ACHILLES was equal to the reported number in 2002, 2010, 2011, 2012 and 2013. The ACHILLES overestimated the number of subjects in the other years. The numbers of annual subjects estimated by the ACHILLES were compared with the previous reported numbers. The numbers of infants match each other except 2009. The difference in the number of infants was only four in 2009.

The ACHILLES calculates and visualizes the statistics for the population-level data from Korean nationwide cohort. Since the

ACHILLES visualizes statistics for diagnosis, drug, procedures and hospital visits by calculating prevalence and frequency per person by using the size and color of the boxes in the tree maps. Users easily assess the statistics for population-level data from Korean nationwide cohort.

Graphs from the ACHILLES about visiting emergency room (ER) and use of cimetidine were depicted in figure 3 and figure 4, respectively. The total number of people visiting ER during whole cohort period was 188,954 with a rising trend. One of the most commonly used drug was cimetidine. Number of subjects using cimetidine was increased from 2002 to 2007, and then decreased. The most commonly used population was women in their 60s and 70s.

Table 2– Comparison between previously reported numbers of all subjects and infants in the cohort and the results from ACHILLES

Year	Number of subjects in cohort		Number of infants aged 0 in the cohort	
	Reported	ACHILLES	Reported	ACHILLES
2002	1.025M	1.025M	9565	9565
2003	1.017M	1.032M	9437	9437
2004	1.016M	1.035M	9320	9320
2005	1.016M	1.036M	8557	8557
2006	1.002M	1.037M	7872	7872
2007	1.020M	1.039M	9766	9766
2008	1.000M	1.027M	9393	9393
2009	0.998M	1.024M	8616	8604
2010	1.002M	1.002M	9032	9032
2011	1.006M	1.006M	9694	9694
2012	1.011M	1.011M	9851	9851
2013	1.014M	1.014M	8825	8825

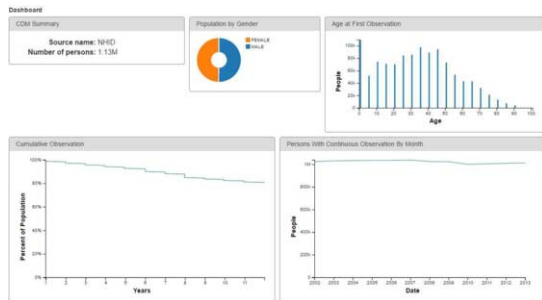


Figure 2– Dashboard of ACHILLES describes the total number of subjects in cohorts according to their gender, race, and year of birth.

Results from ACHILLES Heel

The ACHILLES Heel issued 13 errors and 15 warnings from converted the NHIS-NSC OMOP-CDM data. Most of errors and warnings were resulted from discordance between range of observation period and real date of medical claims

Discussion

We demonstrate successful conversion process of the NHIS-NSC, Korean nationwide cohort database, to the OMOP-CDM version 5.0 in this paper. To date, this has been the first attempt to convert nationwide cohort database to universal standardized

OMOP-CDM format in Asian countries. Although the randomized clinical trial (RCT) undoubtedly remains as a gold standard for developing medical evidence, applicability of evidenced produced by RCTs can be restricted because of gaps between environments in RCTs and real world routine practice [7]. Geographic and racial variations in the risks for disease or the results of treatment also exist. However, these differences have been often neglected in multinational RCTs [8]. Incorporation of Korean longitudinal nationwide large cohort data into the OHDSI network will be the first milestone to generate global high-quality medical evidence, which is applicable to Asian population and real-world practice.

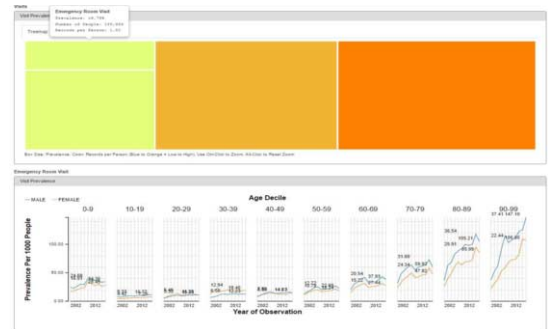


Figure 3– Visit tab of ACHILLES depicts the trend in emergency room visits of subjects in the cohort according to their gender and age.



Figure 4– Drug exposure tab of ACHILLES shows the trend in usage of cimetidine in the cohort according to their gender and age.

Some limitations still exist in the process and the result of our data conversion. Despite regular process in code mapping as previously described [9], information loss was inevitable in the process of code mapping and ETLs owing to unapproved drugs or procedures in US and Korean traditional medical diagnosis and drug. Some tables in the OMOP-CDM, especially regarding to medical cost, were left for the conversion because we planned to construct data set for drug safety monitoring first. Errors issued from the ACHILLES Heel have not fully

investigated. However, the number of error was lower than the median number of 19, which revealed by multi-site evaluation of data quality by using the ACHILLES Heel [5].

Conclusions

We report the successful transformation of the Korean nationwide cohort database, the National Health Insurance Service -National Sample Cohort (NHIS-NSC), into the OMOP-CDM model with acceptable loss of information. Converting additional information including cost data and further verification of OMOP-CDM by replicating previous research are now under way. The OMOP-CDM version of the NHIS-NSC can be a valuable resource for multiple aspects of medical research by incorporation into the OHDSI research network.

Acknowledgements

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI16C0992) and supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI14C3201).

References

- [1] Murdoch TB, Detsky AS. The inevitable application of big data to health care. *The Journal of American Medical Association* 2013;309(13):1351-1352.
- [2] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Studies in Health Technology and Informatics*. 2015;216:574-578.
- [3] Platt R, Wilson M, Chan, KA, Benner JS, Marchibroda J, McClellan M, The new sentinel network — improving the evidence of medical-product safety. *The New England journal of medicine*. 2009 Aug;361(7):645-647.
- [4] Kwon S, Thirty years of national health insurance in South Korea: lessons for achieving universal health care coverage. *Health Policy Plan*. 2009 Jan;24(1):63-71.
- [5] Huser V, DeFalco FJ, Schuemie M, Ryan PB. Multisite evaluation of a data quality tool for patient-level clinical datasets. *eGEMs*. 2016 Nov;4(1).
- [6] Lee J, Lee JS, Park SH, Shin SA, Kim KW. Cohort profile: the National Health Insurance Service-National Sample Cohort (NHIS-NSC), South Korea. *International journal of epidemiology*. 2016 Jan 28;pii:dyv319.
- [7] Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence - What is it and what can it tell us?," *The New England Journal of Medicine*. 2016 Dec 8;375(23):2293-2297.
- [8] Yusuf S, Wittes J. Interpreting geographic variations in results of randomized, controlled trials. *The New England Journal of Medicine*. 2016 Dec 8;375(23):2263-2271.
- [9] Yoon D, Ahn EK, Park MY, Cho SY, Ryan PR, Schuemie MJ, et al. Conversion and data quality assessment of electronic health record data at a Korean tertiary teaching hospital to a Common Data Model for Distributed Network

Research. *Healthcare Informatics Research*. 2016 Jan 31;22(1):54-58.

Address for correspondence

Rae Woong Park, MD, PhD

Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea

164, World cup-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, Korea
Tel: +82-31-219-4470

Fax: +82-219-4472

E-mail: veritas@ajou.ac.kr

Assessing Data Integration and Quality for the Evaluation of Point-of-Care Testing Across Rural and Remote Emergency Departments in Australia

Ling Li^a, Euan McCaughey^a, Juliana Iles-Mann^b, Andrew Sargeant^c, Maria R. Dahm^a,
Virginia Mumford^a, Johanna I Westbrook^a, Andrew Georgiou^a

a Centre for Health Systems and Safety Research, Australian Institute of Health Innovation,
Macquarie University, Sydney, New South Wales, Australia

b Pathology West, New South Wales Health Pathology, Sydney, New South Wales, Australia

c New South Wales Health Pathology, Sydney, New South Wales, Australia

Abstract

In Australia, New South Wales Health Pathology's implementation of managed Point-of-Care Testing (PoCT) services across rural and remote emergency departments (EDs) has the potential to significantly improve access to results for certain types of pathology laboratory tests and help to deliver timely patient care. The aim of this study was to assess the quality of the datasets, including the integration of PoCT results into clinical systems, as a precursor to the application of an evaluation framework for monitoring the delivery of PoCT services and their impact on patient care. Three datasets, including laboratory, ED presentations and hospital admissions data were extracted from the relevant clinical information systems. Each dataset was assessed on six dimensions: completeness, uniqueness, timeliness, validity, accuracy, and consistency. Data incompleteness was the largest problem. Assessing the PoCT data integration and data quality is a precondition for the evaluation of PoCT and for monitoring and improving service delivery.

Keywords:

Point-of-Care Testing; Data Accuracy; Australia

Introduction

Point-of-Care Testing (PoCT) refers to pathology laboratory tests performed near patients and outside a traditional laboratory [1]. PoCT can be conducted in a variety of contexts within the community by patients themselves (typically in their own homes), or in medical environments (hospital bed side, general practice or pharmacies) by clinical personnel who are not necessarily trained in laboratory sciences [1]. The types of tests available range from a consumer-friendly dip stick or prick tests (e.g. home pregnancy tests or glucose meters) to moderately complex (often cartridge based) devices used by trained clinical personnel, to highly sophisticated instruments that can only analyse specifically prepared specimens [2].

Existing evidence shows that the key advantages of a PoCT service are greater access to laboratory testing, especially in regional and remote areas [3; 4], and faster test result turnaround times (TAT) expediting the efficiency of clinical decision making and treatment [5-8]. PoCT has been heralded as being 'ideally suited to [time-critical] circumstances [involving] a small number of analytes (single analyte or single cartridge)' (p.3)[2]. It is thus considered to be particularly valuable for emergency departments (EDs) in helping to meet health professionals' demands for rapid test results which can expedite patient flow and prevent overcrowding [7]. The

majority of PoCT studies have been conducted in urban (most often teaching) hospitals which have regular access to a formal clinical laboratory [9; 10]. However, as a number of researchers have highlighted, the prospective benefits commonly attributed to PoCT (faster TAT and treatment onset times and shorter length of stay [LOS]) are very likely to be evidenced in underserved rural communities [1; 11; 12].

Traditionally, hospitals in rural and remote areas suffer from the 'tyranny of distance' and without on-site laboratory support face extended wait times for laboratory results alongside difficulties in specimen collection and transport [11; 12]. The introduction of PoCT in rural community based health services and hospitals has led to almost immediate access to results, enhanced clinical decision making, faster treatment onset, and disposition to dedicated wards reducing mortality rates and achieving optimal health outcomes [4; 11-13]. These benefits and the potential for improved outcomes rely on careful planning, readily defined roles of stakeholders and a model of clinical care that has been adapted to cater for successful integration of PoCT services [12].

Australia is the third least densely populated country in the world (less than 2.9 people/km²). With approximately one-third of the population living outside major cities, it has one of the lowest population densities outside its major cities [14]. NSW is the most populous state (6.9m) in Australia with more than 800,000 km² land [15]. NSW Health Pathology (NSWHP) is implementing PoCT services across rural and remote EDs, including to areas with extremely limited access to health care services [1; 4; 11]. By the end of 2015, almost 400 PoCT devices had been delivered to 175 EDs in non-metropolitan areas of NSW. To our knowledge this is one of the world's largest managed PoCT services [4; 16]. The rollout of this PoCT service offers the scope for a systematic investigation of the impact of PoCT implementation in rural and remote EDs to explore the operational impacts, evaluate patient outcomes and cost benefits and to develop an evaluation framework to aid PoCT expansion into additional health services such as ambulances and home care.

NSWHP's commitment to the evaluation of PoCT services stems from the need to monitor and enhance the design, implementation and sustainability of the service and ensure the achievement of value for money, the delivery of improved efficiency, effectiveness and optimal patient outcomes. Undertaking an evaluation of PoCT services across rural and remote EDs in NSW involves the examination of data availability and quality. The Royal College of Pathologists of Australasia's (RCPA) PoCT Quality Framework in 2014

recommended that an information management system must be developed to manage information generated by, entered into and transmitted from the PoCT devices to the Laboratory Information System (LIS) and then electronic Medical Records (eMR) systems [17]. It is also one of the key objectives of NSWHP's implementation. The aim of this study is to assess the quality of the datasets, including the integration of PoCT results into clinical systems, as a precursor to develop and apply a robust PoCT services evaluation framework.

Methods

Study design and setting

This was a retrospective before and after cohort study using laboratory and emergency department (ED) data. The study period was from January 2012 to the April 2015. The post implementation period started from January 2014. The evaluation was conducted across EDs in three Local Health Districts (LHDs) in NSW; Far West, Murrumbidgee and Western NSW LHDs (Figure 1). A total of 68 EDs were included in this study (seven in Far West, 26 in Murrumbidgee and 35 in Western NSW).

Laboratory services are provided by Pathology West NSW to hospitals and EDs in these LHDs. The PoCT implementation included the delivery of devices to many EDs that do not have support of a 24/7 laboratory onsite, based on a test profile of 1) Electrolytes, 2) Urea and Creatinine (EUC), 3) Blood gases + lactate + haemoglobin, 4) Troponin and 5) International Normalized Ratio (INR)/ Prothrombin Time (PT).



Figure 1: Local health districts, New South Wales, Australia [18]

PoCT data reconciliation and integration

The PoCT management system 'AQUIRE' stores data relating to every PoCT test ordered at the study EDs. PoCT tests in AQUIRE can be included in the LIS data but only if the patient demographic data entered into the PoCT device at the time of testing can be matched to an individual patient by the AQUIRE middleware which receives this information directly from the Hospital Patient Administration System (HPAS). This process is referred to as data reconciliation for this project. If reconciled, the PoCT results will be integrated into the LIS (Figure 2) and then into the patient's eMR. The AQUIRE data in the LIS was available from the 1st January 2014 after the PoCT implementation.

Data sources and linkage

Three datasets were extracted from clinical information systems. Laboratory data, including the reconciled PoCT data

from AQUIRE, were extracted from the LIS, while ED and inpatient data were extracted from two information systems: Emergency Department Information System (EDIS) and HPAS. Three datasets were linked using a unique and non-identifiable patient Medical Record Number (MRN), as well as their gender and age. This process is depicted in Figure 2. The shaded area shows the ED presentations with laboratory testing including PoCT.

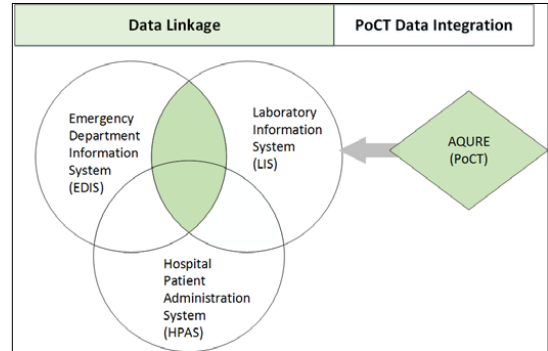


Figure 2: PoCT data integration and data linkage

Data quality assessment

Each dataset was assessed across six dimensions: completeness, uniqueness, timeliness, validity, accuracy, and consistency (Table 1). Consistency was assessed in the linked data across three data sources. The issue of completeness was also examined for the rate of reconciliation for PoCT tests from AQUIRE that were verified through a match with an individual ED presentation and uploaded to LIS.

Table 1: Data quality dimensions [19; 20]

Dimension	Definition
Completeness	The proportion of stored data against the potential of "100% complete"
Uniqueness	Nothing will be recorded more than once based upon how that thing is identified
Timeliness	The degree to which data represent reality from the required point in time.
Validity	Data are valid if it conforms to the syntax (format, type, range) of its definition.
Accuracy	The degree to which data correctly describes the "real world" object or event being described.
Consistency	How well data agree across different data sets, and the extent of agreement between different data sets that are measuring the same thing

Ethics approval

Ethical approval for this study was granted by the Greater Western Area Health Service Human Research Ethics Committee (LNR/15/GWAHS/26). Site Specific Assessment approval to conduct research within each of the LHDs was provided by the Far West LHD (LNRSSA/15/GWAHS/48), Murrumbidgee LHD (LNRSSA/15/MLHD/8) and Western NSW LHD (LNRSSA/15/GWAHS/49).

Results

During the study period, there were 570,538 ED presentations, 262,806 inpatient admissions, and 727,168 laboratory tests, including PoCT tests, across three LHDs.

Completeness

Laboratory data including PoCT data integration

In Far West LHD, only one of the seven EDs had PoCT results reconciled and integrated into the LIS and eMR. In Murrumbidgee, data were provided for all 26 EDs. In Western NSW LHD, limited PoCT results were reconciled and integrated into the LIS and eMR at five EDs.

When comparing the LIS dataset with information from AQUIRE it was found that the LIS data contained fewer patients having PoCTs. Out of 68 EDs, 13 did not have compliant systems resulting in a proportion of PoCT tests that were unable to be reconciled. The average reconciliation rate across 55 EDs over the period encompassing the two to 13 months post the implementation of PoCT was 28.6%, increasing from 20.2% in the second month post the implementation of PoCT to 48.6% in the 13th month post PoCT.

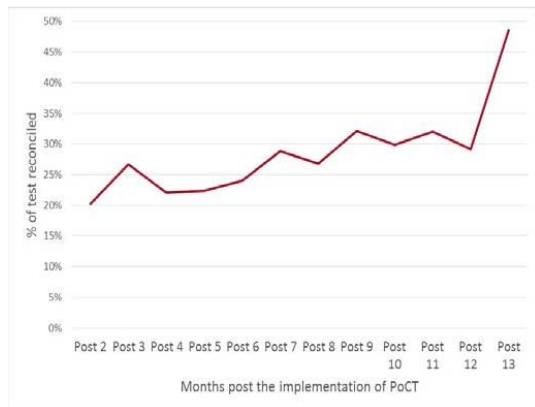


Figure 3: Reconciliation rate of PoCT ordered across 55 EDs in the 2 to 13 months post the implementation of PoCT

ED data

For Far West LHD, the extract only included one ED, with no data for the other six EDs in this LHD. Due to a lack of systems compliant with HIE, ED data for Western NSW LHD were limited to 11 EDs with complete data available from December 2014 onwards.

Murrumbidgee was the only LHD which confirmed the availability of all ED data (Figure 4). However, the number of patient presentations per month in 2012 (range from 1968 to 2597) was much lower than those after January 2013 (range from 4647 to 6293). The median length of stay per month in 2012 (range from 85 to 114 minutes) was much higher than that afterwards (range from 70 to 85 minutes). In terms of the completeness of the data fields, Murrumbidgee was the only LHD that provided a variable distinguishing between 'planned' and 'unplanned' ED presentations.

Inpatient data

The inpatient data extracted from HPAS was confirmed as containing information relating to all inpatient stays in Murrumbidgee. In Far West LHD, the inpatient data were again limited to one site. In Western NSW LHD inpatient data were available for 33 out of 35 sites in this LHD.

Uniqueness

No duplications were identified in the datasets.

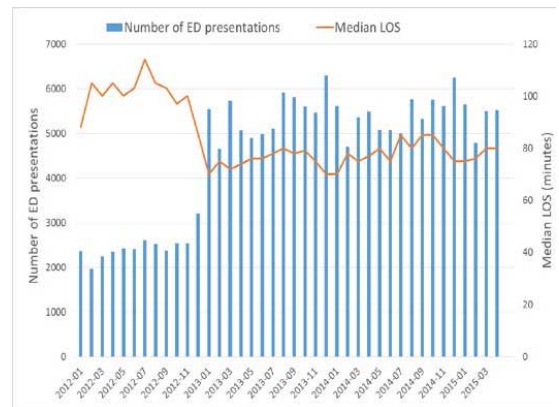


Figure 4: Number of ED presentations and Median LOS in Murrumbidgee LHD

Timeliness

Laboratory data related to laboratory tests up to May 2015, while ED and inpatient data related to presentations/episodes of care up to April 2015. Analysis was performed between September and December 2015. Therefore, the data were regarded as adequate.

Construct validity is the extent to which data measure what they claim to be measuring. It was found that overlapping data in the ED and Laboratory data, such as patient MRN, gender, date of birth and patient location were in agreement. No overlap between data that was not supposed to relate to each other was found.

Accuracy

To determine the accuracy of the data, the contents of all fields were analyzed and suspect entries identified. Such entries were further analyzed by the research team and, after liaison with Pathology West and the LHDs, data deemed to be inaccurate were removed from the dataset.

Consistency

Consistency is particularly pertinent to this analysis, as a high level of consistency is required to enable triangulation of data from different sources through data linkage (Figure 2). It was found that after the accuracy of the data had been established matching data fields in the ED and Pathology data, such as patient MRN, gender, date of birth, were in agreement. Therefore, the data were regarded as consistent.

Discussion

High quality data are required for accurately evaluating the impact of health interventions. Through data quality assessment, limitations and deficiency of datasets can be identified. Evaluation of data quality revealed a number of limitations related to the completeness of available data. Murrumbidgee was the only LHD that was able to supply all ED presentations for the period January 2013 to April 2015. The Western NSW LHD was only able to provide ED data from 10 of 35 sites. However, the completeness of these data could not be verified by the LHD. In Far West LHD data from only one hospital, only one ED was confirmed as containing all ED presentations over the entire study duration. Murrumbidgee

LHD was the only data source which allowed the distinction between ‘planned’ (arranged in advance) and ‘unplanned’ (emergency) ED presentations. This distinction can have an important effect on data analyses and interpretation.

Data linkage provides numerous opportunities to extend our understanding of health care phenomena. Nevertheless, as the challenges of integrating PoCT into hospital networks attest, linked data needs to be carefully assessed for quality. Ensuring the quality of linked data should therefore incorporate data profiling techniques to examine the quality of each dataset separately [20; 21]. This could involve the application of algorithms that identify missing data, duplicates, data formatting and compliance with logic rules (e.g. patient was admitted before they were discharged). It should also incorporate an interrogation of key variables using descriptive statistics to examine the range of findings, percentiles and outliers for consistency and validity [22].

The process of data linkage using hospital, patient and laboratory data involves key ethical, privacy and confidentiality issues involving data governance processes and approval from the appropriate Human Research Ethics Committees (HREC). It also involves controls to protect the integrity of the data and ensure that no identifiable data is publicly available.

Assessing data integration and data quality provided the basis for targeted recommendations for improvement. First, data incompleteness is the largest problem in these rural and remote LHDs. This problem is universal across three clinical information systems: the LIS, EDIS and HPAS. LIS data completeness is also dependent upon the reconciliation of data from HPAS with patient information entered on the PoCT device at the time of testing. Secondly, the reconciliation rate of available PoCT data were only 28.9% although it has been increased from 20.2% in the second month post the implementation of PoCT to 48.6% in the 13th month post PoCT. One of the key recommendation from Royal College of Pathologists of Australasia for quality use of PoCT include the capacity for seamless and automated transfer of results to patients’ electronic medical records [17]. Lack of electronic medical record systems in these rural and remote EDs could be one of main reasons for the low reconciliation rate, in addition to 1) incorrect entry of patient identifications into PoCT devices and 2) lack of resources to ensure that MRNs are available for new patients or that the patient visit is updated in HPAS at time of PoCT.

Conclusion

The NSWHP implemented one of the world’s largest PoCT services across rural and remote EDs. It has the potential to significantly improve access to on-the-spot results for certain types of laboratory tests as a means to deliver timely patient care. Assessing PoCT data integration and data quality is a key precondition for evaluating the implementation of PoCT services and for further monitoring and improving the services into the future. Although our study was based on Australian rural and remote sites, the data integration and quality issue is not unique to these Australian sites, but of relevance to researchers and implementation teams globally.

This study not only provided a rigorous assessment methodology but also highlighted the value of linking routinely collected datasets from different clinical information systems in data quality assessment and subsequently in evaluation of the PoCT implementation.

Acknowledgements

This study was commissioned by NSW Health Pathology and was conducted as part of a project to evaluate the implementation of PoCT service across NSW rural and remote EDs.

AS and JI are employed at NSW Health Pathology. They contributed to the study design and revisions of the manuscript, but were not involved in the data assessment and interpretation. We acknowledge the contribution of Jill Reymont, Dominic Dwyer, Roger Wilson, Keira Robinson, Michael Whiley and Sue Carter to this project.

References

- [1] J.H. Nichols, Point-of-care testing, in: *The Immunoassay Handbook: Theory and applications of ligand binding, ELISA and related techniques*, D. Wild, ed., Elsevier Oxford, UK, 2013, pp. 455-463.
- [2] Royal College of Pathologists of Australasia, Point of care testing, *Common Sense Pathology* April 2015 (2015), 1-8.
- [3] S. Glover and B.V. Bajorek, Exploring point-of-care testing of capillary blood in warfarin management, *Journal of Pharmacy Practice and Research* **38** (2008), 300-304.
- [4] M. Shephard, Point-of-Care testing in Australia: The status, practical advantages, and benefits of community resiliency, *Point of Care* **12** (2013), 41-45.
- [5] E. Lee-Lewandrowski and K. Lewandrowski, Point-of-care testing in the Emergency Department, in: *Point-of-Care Testing: Needs, Opportunity and Innovation*, C. Price, A. St John, and L.L. Kricka, eds., AACC Press, Washington, DC, 2010, pp. 397-410.
- [6] V. Pecoraro, L. Germagnoli, and G. Banfi, Point-of-care testing: where is the evidence? A systematic survey, *Clinical Chemistry and Laboratory Medicine* **52** (2014), 313-324.
- [7] K.D. Rooney and U.M. Schilling, Point-of-care testing in the overcrowded emergency department - Can it make a difference?, *Critical Care* **18** (2014).
- [8] L. Li, A. Georgiou, E. Vecellio, A. Eigenstetter, G. Toouli, R. Wilson, and W. JI, The impact of pathology testing on Emergency Department length of stay: A multi-hospital longitudinal study applying a cross-classified random effect modeling approach, *Academic Emergency Medicine* **22** (2015), 338-346.
- [9] M. Bradburn, S.W. Goodacre, P. Fitzgerald, T. Coats, A. Gray, T. Hassan, J. Humphrey, J. Kendall, J. Smith, and P. Collinson, Interhospital variation in the RATPAC trial (Randomised Assessment of Treatment using Panel Assay of Cardiac markers), *Emergency Medicine Journal* **29** (2012), 233-238.
- [10] P. Collinson, S. Goodacre, D. Gaze, and A. Gray, Very early diagnosis of chest pain by point-of-care testing: comparison of the diagnostic efficiency of a panel of cardiac biomarkers compared with troponin measurement alone in the RATPAC trial, *Heart* **98** (2012), 312-318.
- [11] K. Blattner, G. Nixon, C. Jaye, and S. Dovey, Introducing point-of-care testing into a rural hospital setting: thematic analysis of interviews with providers, *J Prim Health Care* **2** (2010), 54-60.
- [12] M. Shephard, R. Tirimacco, and P. Tideman, Point-of-Care testing in remote environments, in: *Point-of-Care Testing: Needs, Opportunity and Innovation*, C. Price, A. St John, and L.L. Kricka, eds., AACC Press, Washington, DC, 2010, pp. 373-386.

- [13] P.A. Tideman, R. Tirimacco, D.P. Senior, J.J. Setchell, L.T. Huynh, R. Tavella, P.E.G. Aylward, and D.P.B. Chew, Impact of a regionalised clinical cardiac support network on mortality among rural patients with myocardial infarction, *Medical Journal of Australia* **200** (2014), 157-160.
- [14] Australian Bureau of Statistics (ABS), Outcomes of ABS Views on Remoteness Consultation, Australia, Jun 2001, in, 2016.
- [15] Australian Bureau of Statistics (ABS), Population by Age and Sex, Regions of Australia, 2011 Census (3235.0), in, 2012.
- [16] NSW Health Pathology, Point of care testing, in, 2015.
- [17] Royal College of Pathologists of Australasia (RCPA), *Point of Care testing: Elements of a quality framework*, RCPA, Surry Hills, 2014.
- [18] Head Medical, New South Wales Health, in.
- [19] Data Quality Dimensions Working Group, *The Six Dimensions of EHDI Data Quality Assessment*, DAMA UK, 2013.
- [20] T.L. Strome, *Healthcare Analytics for Quality and Performance Improvement*, John Wiley & Sons, Inc, Hoboken, NJ, USA, 2013.
- [21] T.C. Redman and M. Daugherty, *Data Quality: The Field Guide*, Digital Press, Greenwich, CT, USA, 2000.
- [22] R.C. Hawkins, Laboratory turnaround time, *Clinical Biochemist Reviews* **28** (2007), 179-194.

Address for correspondence

Corresponding author: Ling Li

Email: ling.li@mq.edu.au

Full mailing address: Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia

Tel: +61 2 9850 2423

Fax: +61 2 9850 2499

Using Machine Learning Models to Predict In-Hospital Mortality for ST-Elevation Myocardial Infarction Patients

Xiang Li^a, Haifeng Liu^a, Jingang Yang^b, Guotong Xie^a, Meilin Xu^c, Yuejin Yang^b

^a IBM Research – China, Beijing, China

^b Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China

^c Pfizer Investment Co. Ltd., Beijing, China

Abstract

Acute myocardial infarction is a major cause of hospitalization and mortality in China, where ST-elevation myocardial infarction (STEMI) is more severe and has a higher mortality rate. Accurate and interpretable prediction of in-hospital mortality is critical for STEMI patient clinical decision making. In this study, we used interpretable machine learning approaches to build in-hospital mortality prediction models for STEMI patients from Chinese Acute Myocardial Infarction (CAMI) registry data. We first performed cohort construction and feature engineering on CAMI data to generate an available dataset and identify potential predictors. Then several supervised learning methods with good interpretability, including generalized linear models, decision tree models, and Bayes models, were applied to build prediction models. The experimental results show that our models achieve higher prediction performance (AUC = 0.80–0.85) than the previous in-hospital mortality prediction STEMI models and are also easily interpretable for clinical decision support.

Keywords:

Myocardial Infarction; Hospital Mortality; Machine Learning

Introduction

Acute myocardial infarction (AMI) is a major cause of hospitalization and mortality in China, where the in-hospital mortality rate for ST-elevation myocardial infarction (STEMI) is even higher than that for non-ST-elevation myocardial infarction (NSTEMI) [1,2]. Because there is considerable variability in mortality risk among patients with STEMI, it is critical to accurately predict the risks of in-hospital mortality for STEMI patients at the time of hospital presentation, in order to decide on the allocation of clinical resources and the choice of interventional and medical therapies [3,4].

Current in-hospital mortality risk models for STEMI, such as TIMI score [3] and GRACE score [4], use predictors that are ground in previous known evidence, including age, heart rate, systolic blood pressure (SBP), Killip levels, weight, history of hypertension, diabetes and angina, anterior STEMI, time to treatment, serum creatinine, and cardiac arrest. These risk scores were derived from logistic regression models, which are well understood and easy to apply in clinical practice. However, the performance of in-hospital mortality prediction for STEMI still has room for improvement. Besides the predictors used in the existing models, there are other potential predictors that are highly related to STEMI in-hospital mortality, which can also be used in risk prediction to improve the performance. Besides, some other machine learning models may achieve better prediction performance or interpretability than regression analysis models like logistic regression.

Therefore, this study investigated the modeling of in-hospital mortality prediction models in STEMI that have good prediction ability and interpretability using machine learning methods. The study was based on the Chinese Acute Myocardial Infarction (CAMI) [2] data, which collected the patients' demographics, symptoms, medical history, results of physical examination and laboratory test, details of in-hospital treatments, and clinical events including mortality. Because the dataset is heterogeneous and redundant, dimensionality reduction methods and model learning algorithms that can handle redundant feature sets should be used for predictive modeling. Moreover, the objective of this study was to build human understandable and applicable risk prediction models. Though many previous works used feature engineering and supervised learning methods to build high accuracy risk prediction models for cardiovascular diseases and diabetes [5–10], the learning methods in which resulting models are difficult to interpret (e.g., principle component analysis, support vector machine, and deep neural network) are not preferable in building clinically interpretable risk models.

In this study, we integrated interpretable machine learning approaches to build in-hospital mortality risk prediction models for STEMI patients from CAMI data. Feature engineering methods, including feature construction, missing data imputation and filter-based feature selection, were used to generate available dataset, identify potential predictors and reduce redundancy. Then we applied different categories of supervised learning methods that have good interpretability, including generalized linear models, decision trees, and Bayes models, to build risk prediction models that are suitable for different clinical scenarios. The experimental results show that our models can achieve higher prediction performance than the previous risk prediction models for STEMI, and are also easily interpretable for clinical decision support.

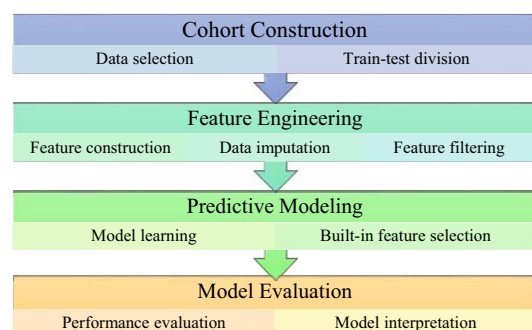


Figure 1 – Pipeline of building in-hospital mortality risk prediction models for STEMI patients

Methods

Figure 1 shows our pipeline of building in-hospital mortality risk prediction models for STEMI patients. We first constructed the cohort of interest, and applied feature engineering to identify potential predictors. Then we trained prediction models using supervised learning algorithms, and evaluated their prediction performance and model interpretability.

Cohort Construction

The dataset used in this study were collected in the CAMI registry project [2]. The project started in 2013 and 26,103 patients with AMI were registered until 2014. From the CAMI data, we identified 18,744 patients who were hospitalized due to STEMI in 2013 and 2014, where 1,263 patients were cases who died in hospital (the in-hospital mortality rate is 6.74%), and the others were control instances who survived in hospital. The features used in this study were those collected at the time of hospital presentation (i.e., before any in-hospital treatment). In the CAMI data, 132 original features meet this criteria, including demographics, medical and treatment histories, life styles, onset symptoms, initial in-hospital vital signs, laboratory test results, etc.

In this study, we used the data of patients hospitalized in 2014 as the training set (9,619 patients, whose mortality rate is 6.78%) to develop the risk prediction models and the data of patients hospitalized in 2013 as the testing set (9,125 patients, whose mortality rate is 6.70%) to valid the prediction performance of the models.

Feature Engineering

In the raw CAMI data, some original features are not appropriate to be directly used in predictive modeling (e.g., birth year), and a large proportion of original features (more than 95%) have missing values. Moreover, not all original features are highly related to in-hospital mortality of STEMI. Therefore, we performed feature engineering to construct and select the potential predictors for risk prediction. We first transformed the original features to features that are easy to analyze by feature construction. For example, the “birth year” of each patient was transformed to “age”. Then, we employed data imputation to fill-in missing values and applied filter-based feature selection algorithms to identify potential predictors from the imputed features.

The raw CAMI data has significant omissions due to the questionnaire structure; unknown values or errors in data collection. Therefore, we performed missing data imputation before predictive analysis. Firstly, the features with too many missing entries (more than 20%) were discarded, because their distributions are difficult to estimate. For the remaining features, every missing value of a numeric feature (e.g., SBP) was replaced with the mean of the feature’s observed values, every missing value of an ordinal feature (e.g., Killip level) was replaced with the median of its observed values, and every missing value of an unordered nominal feature (e.g., history of diabetes) was replaced with the mode of its observed values. In this study, after feature construction and missing data imputation, 93 available candidate features were produced to the following feature selection step.

In machine learning, feature selection methods automatically test and select predictive features from a large number of candidate features. There are three main supervised feature selection strategies: filter, wrapper, and embedded models [11]. The filter models separate feature selection from model learning, and the wrapper and embedded models integrate feature selection in learning process. In the step of feature engineering, we performed filter models to remove the features

that do not provide useful information and select the features that have high relevancy against the outcome. Concretely, the close-to-constant features, in which 99% of the instances have identical values, were first removed. Then we employed and compared two feature filtering models.

- **Univariate filter.** A univariate filter method calculates a score to represent the relevancy of a feature against the outcome, and filter the feature based on the score independently. In this study, we used the p-value from two standard statistical tests, the Chi-square test for categorical features and the ANOVA F test for numeric features, as the relevancy scores and selected the features whose p-value < 0.05.
- **Multivariate filter.** Different from the univariate filter method, a multivariate filter method evaluates the input features as a batch producing a subset of features that have the highest overall score. In this study, we used the correlation-based feature subset selection (CFS) method [12] to obtain the subset of features highly correlated with the outcome while having low intercorrelation between the features.

In this study, we also combined the features that were automatically selected by the filter-based feature selection algorithm with features that are well-known risk factors from prior knowledge [3,4], but were not automatically selected (e.g., anterior STEMI, time to treatment) as the predictors to build risk prediction models.

Predictive Modeling

In this study, we applied and compared different categories of machine learning models that have good interpretability, including generalized linear models (GLM), decision tree models, and Bayes models, to develop in-hospital mortality prediction models for STEMI patients. Besides, built-in feature selection strategies, including wrapper and embedded selection, were employed in some modeling processes.

- **Generalized linear model.** GLM generalizes ordinary linear regression by allowing the linear model to be related to the response variable via a link function, which is widely used in both medical statistics and machine learning due to its good prediction performance and interpretability. In this study, we applied logistic regression (LR), which is a GLM with a logit link function and a binomial distribution, and Cox proportional hazards model [13], which is a semi-parametric GLM that takes into account the time of censoring. We also employed forward stepwise feature selection, which is a wrapper selection model, to evaluate the performance and statistical significance of the LR and Cox models under different selection of features.
- **Decision tree model.** Decision trees are very easy to interpret and therefore have been successfully applied in healthcare. Decision tree learning algorithms usually embed feature selection into the learning process of a model when splitting the source data set into subsets. In this study, we employed the Chi-squared automatic interaction detector (CHAID) [14] method to build tree-based prediction model. CHAID is an efficient statistical technique that uses significance of Chi-squared test as a criterion for tree growing. We also employed random forest [15], which constructs multiple random decision trees and integrate the outputs of the trees for prediction. Compared to single decision tree models like CHAID, random forest reduces the problem of over-fitting, but has worse interpretability.
- **Bayes model.** Bayes models can learn probabilistic relationships among features and an outcome; computing

the probabilities of the outcome given the features. The Bayes models are also interpretable based on Bayes theorem. In this study, we employed naive Bayes [16], which is a Bayes model with strong independence assumptions between the input features and therefore can be trained very efficiently. We also applied Bayes network [17], which is a probabilistic graphical model that represents features and their conditional dependencies via a directed acyclic graph, where both the dependencies between outcome and input features and the interdependencies among input features can be modeled.

Results

We evaluated the performance of our approaches in building in-hospital mortality risk prediction models for STEMI patients from the CAMI dataset. The area under the receiver operating characteristic curve (AUC) were used to evaluate the prediction performance of models. We performed feature engineering and model learning on the training set, applied the learned models on the testing set, and evaluated the AUC on both datasets.

We first generated four different feature sets by feature engineering:

1. None of the feature filtering methods were applied, all original 93 features were kept.

2. Univariate filter selection was performed to select 51 features
3. CFS, which is a multivariate filter algorithm, selected 19 features
4. A combination of features from CFS and prior knowledge [3,4].

Then we built different learning models described above using different feature sets and evaluated their prediction performance. The results are shown in Table 1, where the models with less features and higher AUC on the testing set are highlighted. Random forest and the Bayes network achieved the best performance when applied on all candidate features, but their performance cannot be increased by filter-based feature selection. In comparison, after performing filter-based feature selection, the performance of GLM methods (LR and Cox) increased. Moreover, the combination of auto-selected features by CFS and prior knowledge based features improved the prediction performance for the majority of learning models.

We also compared the performance of our approaches to the state-of-the-arts risk models: TIMI score [3] and GRACE score [4]. We applied both previous models and our trained models on the same testing set and computed the AUC. As shown in Figure 4, the prediction performance of most of our models outweighed the TIMI and GRACE models.

Table 1 – AUC of different learning models on different feature sets

Feature selection	1) None			2) Univariate filter			3) CFS			4) Combination		
	No. feature	AUC (train)	AUC (test)	No. feature	AUC (train)	AUC (test)	No. feature	AUC (train)	AUC (test)	No. feature	AUC (train)	AUC (test)
LR	93	0.858	0.836	51	0.852	0.842	19	0.840	0.841	26	0.844	0.843
LR stepwise	21	0.846	0.839	19	0.845	0.839	15	0.840	0.840	17	0.842	0.843
Cox	93	0.853	0.829	51	0.849	0.838	19	0.839	0.840	26	0.842	0.842
Cox stepwise	21	0.842	0.835	18	0.843	0.835	14	0.838	0.838	16	0.840	0.839
CHAID	11	0.818	0.796	11	0.818	0.794	7	0.811	0.801	7	0.811	0.801
Random forest	93	0.917	0.849	51	0.915	0.842	19	0.898	0.843	26	0.901	0.846
Naive Bayes	93	0.820	0.818	51	0.820	0.823	19	0.817	0.820	26	0.821	0.825
Bayes network	93	0.872	0.846	51	0.867	0.840	19	0.865	0.835	26	0.868	0.835

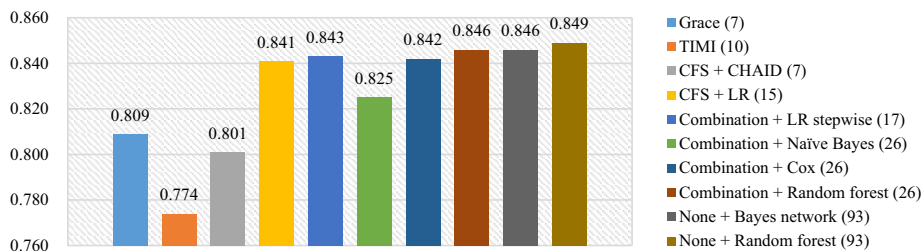


Figure 2 – AUC of different models, evaluated on the same testing set.

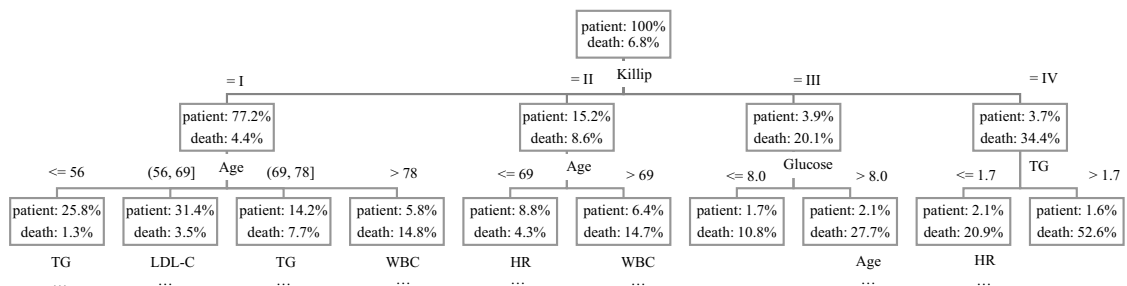


Figure 3 – CHAID tree model on CFS feature set

Since the objective of this work was to develop in-hospital mortality prediction models that can be used in real world clinical practices, the interpretability of the models was as important as the prediction performance. Table 2 shows the stepwise LR model built on the Combination feature set (AUC: 0.842, 95% confidence interval (CI):0.826-0.859). As a traditional regression analysis model, the contribution of each selected feature can be represented by the odds ratio (OR), and its statistical significance can be evaluated using 95% CI and p-value. For example, in the LR model of Table 2, for every 10-year increase in age the odds of in-hospital mortality multiplies by 1.654. The cox regression models are similar for interpretation, where the contribution of each feature can be represented by the hazard ratio.

Table 2 – Stepwise LR Model on Combination feature set

Feature	OR	95% CI		p
History of CABG	5.278	1.482	18.795	0.010
Cardiac shock	2.346	1.631	3.374	<0.001
Killip = I (referent)				<0.001
Killip = II	0.892	0.679	1.173	0.414
Killip = III	1.449	1.012	2.074	0.043
Killip = IV	1.910	1.317	2.770	0.001
Atypical presentation	1.776	1.305	2.418	<0.001
Age (per 10)	1.654	1.515	1.807	<0.001
Malignant arrhythmia	1.496	1.137	1.969	0.004
Anterior STE	1.346	1.111	1.631	0.002
Heart failure	1.326	1.019	1.727	0.036
Heart rate (per 10)	1.224	1.172	1.277	<0.001
Potassium (per 1)	1.210	1.037	1.413	0.016
WBC (per 10 ⁹)	1.083	1.058	1.109	<0.001
Glucose (per 1)	1.051	1.026	1.075	<0.001
Creatinine (per 10)	1.039	1.025	1.053	<0.001
Weight (per 10)	0.888	0.804	0.982	0.020
SBP (per 10)	0.872	0.839	0.907	<0.001
Living with spouse	0.743	0.604	0.913	0.005
Sex (male)	0.666	0.539	0.821	<0.001

Though the AUC of the CHAID tree models are not as good as our other models, the CHAID model has the very clear interpretation. As shown in Figure 3, in the CHAID model built on the CFS feature set, the whole dataset can be split into four subgroup with very different mortality rates by Killip level. Also, the patient group with Killip = I can be further divided into four smaller subgroups by age. Therefore, each patient belong to a leaf node in the tree that can be uniquely defined using a set of rules, and the mortality rate of this node (subgroup) is explicitly used to predict the patient's risk. In contrast, though random forest achieved the best prediction performance on every feature set of this study, each random forest model has many different decision trees (100 trees in our setting) and is not easy to interpret.

A Bayes model can represent the conditional dependencies between input features and outcome, and also has good interpretability. Because the interdependencies between features are complex in the Bayes network models built in previous experiments, for demonstration purpose we show the Bayes network model developed on the 12 well-known predictors using our training dataset in Figure 4. Both the dependencies between the outcome and the predictors (e.g., the death outcome has strong direct dependencies on creatinine, Killip level, heart rate, SBP and hypertension) as well as the interdependencies between the predictors (e.g., dependency between heart rate and Killip level, dependency between SBP and hypertension, etc.) are clearly represented in the Bayes network model.

Discussion

In this study, we compared the prediction of several feature

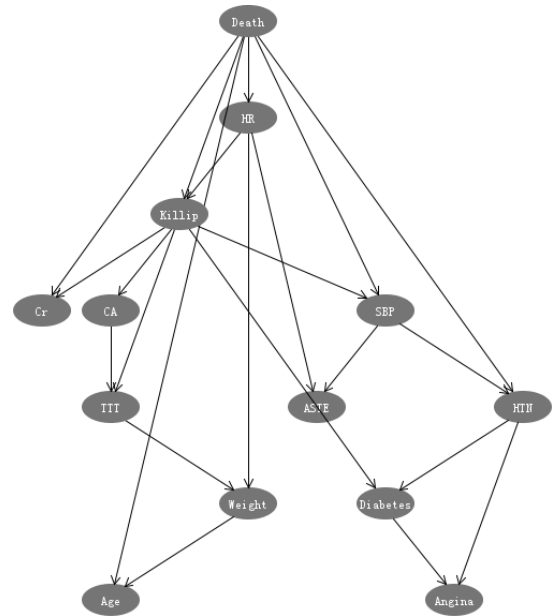


Figure 4 – Bayes network model built on knowledge-based features. HR = heart rate; Cr = creatinine; CA = Cardiac arrest; TTT = time to treatment; ASTE = anterior STEMI; HTN = hypertension

selection and supervised learning methods in building in-hospital mortality prediction models for STEMI patients. For GLM models (LR and Cox), appropriate feature selection can not only reduce the model complexity, but also improves the prediction performance. This is probably because GLM makes the assumption of no multicollinearity between the features, but the whole feature set is redundant, which negatively affects prediction performance. The feature selection methods, including CFS, which minimizes the intercorrelation and stepwise selection that optimizes the performance and statistical significance, can reduce the redundancy of features and therefore increase prediction performance. In contrast, the random forest and Bayes network methods can handle the redundant and intercorrelated features and therefore achieved the best prediction performance on the whole feature set. Moreover, the known predictors from prior knowledge [3,4] were grounded in previous evidence. Adding them to the auto-selected features from our dataset essentially includes the information outside the dataset, and therefore improved the prediction performance.

We also compared the interpretability of different machine learning models for risk prediction. As there is a trade-off between prediction performance and model interpretability, the choice of machine learning models may vary depending on the real-world clinical scenario.

5. Clinicians need to quickly estimate a patient's risk without any decision support tool requiring a human-memorable risk prediction model. Though the prediction performance of a single decision tree model (e.g., the CHAID model, Figure 3) is usually not as good as other machine learning models, it is human understandable and memorable, and is very suitable for this scenario.
6. Clinicians can predict a patient's risk with an independent risk prediction tool, such as a risk calculator, but still needs to manually input the predictor values. For these cases, a model that is

developed by combining feature selection and GLM (e.g., the stepwise LR model in Table 2) can provide decent prediction performance while keeping the input workload acceptable.

7. Clinicians can directly load a patient's data from the health information system (HIS) to perform risk prediction. For these cases, the complex risk prediction models with higher prediction performance (e.g., Bayes network and random forest models built on the whole feature set) can be used in clinical decision support.

For the purpose of interpretability, we did not apply more complex machine learning models such as deep neural network (DNN). However, there already have been attempts to make traditionally uninterpretable models interpretable. For example, Che et al. [18] developed a mimic learning approach, which can derive interpretable decision tree models from DNN models and maintain DNN's strong prediction performance. For future work, we would follow this direction in order to develop interpretable risk prediction models with higher prediction performance.

Another limitation of this work is that we only used a standard data imputation method based on column mean, median, and mode to remedy missing values. Some more advanced statistical imputation methods like multiple imputation, as well as the machine learning based imputation methods such as k-nearest neighbors imputation and neural network imputation, could be tried in the future, to make a more accurate estimation for missing values.

Conclusions

ST-elevation myocardial infarction (STEMI) is a major cause of hospitalization and has high in-hospital mortality rate. Accurate and interpretable prediction of in-hospital mortality is critical for clinical decision making to STEMI patients. In this study, we used integrated machine learning approaches, including the feature engineering and supervised learning methods that have good interpretability, to build in-hospital mortality prediction models for STEMI patients from CAMI data. The experimental results show that our models achieve higher prediction performance than previous models, and are also easily interpretable for clinical decision support.

References

- [1] J. Li, Q. Wang, S. Hu, Y. Wang, F.A. Masoudi, J.A. Spertus, H.M. Krumholz, L. Jiang, China Peace Collaborative Group, ST-segment elevation myocardial infarction in China from 2001 to 2011 (the China PEACE-Retrospective Acute Myocardial Infarction Study): a retrospective analysis of hospital data, *Lancet* **385** (2015):441-51.
- [2] H. Xu, J. Yang, S.D. Wiviott, M.S. Sabatine, E.D. Peterson, Y. Xian, M.T. Roe, W. Zhao, Y. Wang, X. Tang, The China Acute Myocardial Infarction (CAMI) Registry: A national long-term registry-research-education integrated platform for exploring acute myocardial infarction in China, *Am Heart J* **175** (2016):193-201.
- [3] D.A. Morrow, E.M. Antman, A. Charlesworth, R. Cairns, S.A. Murphy, J.A. de Lemos, R.P. Giugliano, C.H. McCabe, E. Braunwald, TIMI risk score for ST-elevation myocardial infarction: A convenient, bedside, clinical score for risk assessment at presentation an intravenous nPA for treatment of infarcting myocardium early II trial substudy, *Circulation* **102** (2000):2031-2037.
- [4] C.B. Granger, R.J. Goldberg, O. Dabbous, K.S. Pieper, K.A. Eagle, C.P. Cannon, F. Van De Werf, A. Avezum, S.G. Goodman, M.D. Flather, K.A. Fox, Predictors of hospital mortality in the global registry of acute coronary events, *Arch Intern Med* **163** (2003):2345-2353.
- [5] A. Khosla, Y. Cao, C.C. Lin, H.K. Chiu, J. Hu, H. Lee, An integrated machine learning approach to stroke prediction, In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010):183-192.
- [6] H. Neuvirth, M. Ozery-Flato, J. Hu, J. Laserson, M.S. Kohn, S. Ebadollahi, M. Rosen-Zvi, Toward personalized care management of patients at risk: the diabetes case study, In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (2011):395-403.
- [7] J. Sun, J. Hu, D. Luo, M. Markatou, F. Wang, S. Ebadollahi, Z. Daar, W.F. Stewart, Combining knowledge and data driven insights for identifying risk factors using electronic health records, In: *proceedings of the American Medical Informatics Association Symposium* **2012** (2012):901-10.
- [8] O. Ogunyemi, D. Kermah, Machine learning approaches for detecting diabetic retinopathy from clinical and public health records, In: *proceedings of the American Medical Informatics Association Symposium* **2105** (2015): 983-90.
- [9] X. Li, H. Liu, X. Du, P. Zhang, G. Hu, G. Xie, Integrated machine learning approaches for predicting ischemic stroke and thromboembolism in atrial fibrillation, In: *proceedings of the American Medical Informatics Association Symposium* **2016** (2016):799-807.
- [10] E. Choi, A. Schuetz, W.F. Stewart, J. Sun, Using recurrent neural network models for early detection of heart failure onset, *J Am Med Inform Assoc* **2016** (2016):799.
- [11] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, *Dat Class Algorithms Appl* (2014):37.
- [12] H.Z. Hamilton, *Correlation-Based Feature Subset Selection for Machine Learning*, New Zealand, 1998.
- [13] D.R. Cox, Regression models and life-tables, *J R Stat Soc B* **34** (1972):187-220.
- [14] G. Kass, An exploratory technique for investigating large quantities of categorical data, *Appl Stat* **29** (1980):119-27.
- [15] L. Breiman, Random forests, *Mach Learn* **45** (2001):5-32.
- [16] G.H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, *Uncertainty in Artificial Intelligence Proc* (1995):338-345.
- [17] J. Pearl, Bayesian networks: a model of self-activated memory for evidential reasoning. Proceedings of the 7th Conference of the, In: *Proceedings of the 7th Conference of the Cognitive Science Society* (1985):329-34.
- [18] Z. Che, S. Purushotham, R. Khemani, Y. Li, Interpretable deep models for ICU outcome prediction, In: *proceedings of the American Medical Informatics Association Symposium* **2016** (2016):371-80.

Address for correspondence

Xiang Li: IBM Research – China, Building 19, Zhongguancun Software Park, Haidian District, Beijing, China, 100193. Email: lixiang@cn.ibm.com

Yuejin Yang: Department of Cardiology, Fuwai Hospital, National Center for Cardiovascular Diseases, Beijing, China, 100037. Email: yangyjf@126.com

Extracting Eligibility Criteria from the Narrative Text of Scientific Research Articles

Ching-Yun Lin^a, Der-Ming Liou^{a,b}, Mei-Lien Pan^c

^a Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

^b Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, Kaohsiung, Taiwan, China

^c Institute of Information Science, Academia Sinica, Taipei, Taiwan, China

Abstract

Eligibility criteria among hundreds of National Health Insurance Research Database (NHIRD) research papers have similar constituent elements, such as demographic characteristics or diagnostic codes. The study results of the same disease could vary among different research due to the variation of the criteria statements, therefore the narrative patterns analysis tool would be helpful for summarizing the knowledge implicitly contained in the eligibility criteria. In this study, we developed a series of R-based text processing methods to extract the narrative eligibility criteria in NHIRD papers by simplifying the article titles and content paragraphs, identifying medical concepts and abbreviations, then detecting basic demographic characteristics and ICD-9-CM diagnosis codes. Although there is still room for improvement on study type identifying, the high performance in classifying the study type, detecting age restrictions and extracting ICD-9-CM codes still shows the system usefulness for the analysis of eligibility criteria.

Keywords:

Database,; Natural Language Processing

Introduction

The de-identified and sampled claim dataset called the National Health Insurance Research Database (NHIRD) has been accessible by researchers purely for research purposes since 2000. This anonymous database contains comprehensive information on insured people, such as demographic data, ambulatory and inpatient medical claims, disease diagnoses, and prescription records. The diagnostic codes used were based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). There are nearly 3,000 NHIRD studies published in international journals, and the number of articles has increased year by year.

Most of NHIRD papers aim to figure out the relationship among disease, medication, and procedure by analyzing the medical records in insurance claim data rather than the development of practical clinical trials. In these NHIRD studies, researchers must define a set of suitable eligibility criteria (also known as “case definition”) of study samples. Just like the clinical trial protocol, the eligibility criteria of NHIRD studies can be roughly divided into inclusion and exclusion criteria [1]. Inclusion criteria are the general description of the target population, while exclusion criteria are further fine-tuned in order to exclude patients that have some characteristics (e.g. medical history, not the age of onset

of the disease) which might interfere with the study results. For example, the narrative of inclusion criteria of chronic kidney disease (CKD) might be “The claims-based diagnosis of CKD was defined by the presence of 1 inpatient or outpatient ICD-9-CM code 585 in the claims and without catastrophic illness registration cards for ESRD” [2], and an exclusion criteria of polycystic ovary syndrome (PCOS) like “Women with PCOS diagnosed at < 15 or > 45 years of age were excluded” [3]. However, there are still differences in the writing patterns. For example, when considering the accuracy of diagnosis, NHIRD researchers emphasize that they use the diagnosis records from specific data files and fields, or the minimum limit of medical visits, rather than physiological values and vital signs which are commonly used in the clinical protocols.

Several research that focus on eligibility criteria have suggested that study results (e.g. prevalence, hazard ratio) of the same disease would vary with a different setting of eligibility criteria. For instance, the incidence rate of Kawasaki disease defined by ICD-9-CM code (446.1) with ATC code (J06BA02) was much lower than defined by ICD-9-CM code (446.1) only [4]; the estimated prevalences of cerebral palsy would differ from 6 kinds of criteria setting [5]; in the study of the risk of epithelial ovarian cancer (EOC) in patients with endometriosis (EM) [6], the authors compared 13 different criteria of EM, the loosest criterion just have 1 outpatient or hospitalization medical record with ICD-9-CM code 617.0-9 while the strictest must be confirmed by code for surgery (65.1X, 65.2X), and it is indicated that the looser the criterion used, the higher the prevalence of EM and the lower the incidence of subsequent EOC would be estimated. It goes to show that stricter eligibility criteria can improve the homogeneity of sample and research accuracy [1], but the number of samples would decrease at the same time [7]. On the contrary, if criteria were too loose, they might lead to overestimation or underestimation of the disease prevalence or even the correlation between disease and its risk factors [6], which would make the study results not representative.

Researchers have to search the ICD-9-CM code or other characteristics of target diseases from previous papers when designing their eligibility criteria. In addition, finding suitable references and then understanding the narrative is a time-consuming labor because researchers tend to use different vocabulary and narratives to avoid the doubts about plagiarism. In nearly 3,000 NHIRD articles, one criterion of the same disease might have dozens of varied statements, causing great obstacles in summarizing the knowledge in eligibility criteria.

If there is a processing mechanism of converting the eligibility criteria into structured form and further summarizing as a retrieval system, it would be useful for a researcher to refer to the existing criteria and facilitate their own study design. In addition, eligibility criteria can further be used as study material, for comparing the results from varied criteria and clarifying the suitable criteria for a specific disease in epidemiological research. There are studies that use natural language processing technologies in criteria text processing that focus on the documents of the clinical trial protocol [8-10] or descriptive clinical text [11; 12] in electronic medical record (EMR), while the criteria of NHIRD articles were written in a much simpler way, thus could be processed by pattern recognition tools without complicated NLP techniques.

In this study, we built a series of text processing tools to extract the narrative eligibility criteria in NHIRD articles. In the following sections, we will describe the methods of retrieving the criteria section in articles, detecting the medical concept, age and gender limitation of subjects, extracting the ICD-9-CM codes, detecting the restriction of specialist, and present the result of system performance.

Methods

Materials

All of our input data arrived from the NHIRD related articles of PLOS ONE journal which provided the open access XML format files (n=280). In the following description, we will introduce the article titled “Increased Risk of Primary Sjögren's Syndrome in Female Patients with Thyroid Disorders: A Longitudinal Population-Based Study in Taiwan” [13] as an example, and using the section titled within curly brackets (e.g. {Abstract}, {Introduction}, {Materials and Methods}) to represent the section of NHIRD articles.

Defining study events

In this study, we only focus on the issues of the diseases. We used the word “event” to represent the medical disease concept in the NHIRD researches. There is a temporal relationship between events in the same article, the order of occurrence event in the example paper is (1) thyroid disorders (2) primary Sjögren's syndrome.

Text processing

Our text processing framework is depicted in Figure 1. The processing works are all developed under R software (v.3.3.0).

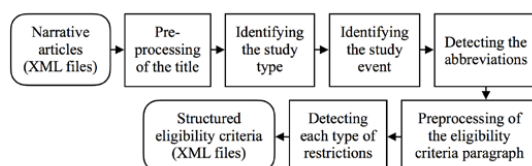


Figure 1 - Text processing methods overview

Preprocessing of the article title

To simplify the article title, the segment sentences unrelated to study event such like “A Longitudinal Population-Based Study in Taiwan” and the words “patient”, “people”, “subject”,

“adult” and “child” would be removed in this step, for avoiding the interference in later event identifying step.

In addition, some special diseases would be written in non-English vocabulary (e.g. “Sjögren” was a Swedish word), that could not be processed by the UMLS mapping tool MetaMap (2014 Linux version) [14], so we converted the Latin alphabets into English alphabets in this step. As the above processing, the title of the example article would be simplified to “Increased Risk of Primary Sjogren's Syndrome in Thyroid Disorders”.

Identifying the study type

In this study, we excluded the articles exploring the issues of disease mortality, socioeconomic factors, cost-effectiveness or utilization of the medical resource. The titles of remaining papers would be filtered with 6 regular expressions (Table 1) to find those that belong to the study type “Increased or decreased (Event 2) risk after (Event 1)”.

Table 1 - Regular expressions of study types

Exclusion
More than two study events <ul style="list-style-type: none"> • (increase_decrease) (...*) (prepositions) (...*) (prepositions) (...*) • (...*) ((increase_decrease) (prepositions) (association_words)) (...*) (prepositions) (...*) • Comparison between (...*) and (...*)
Bidirectional association between two study events <ul style="list-style-type: none"> • Bidirectional (...*) and (...*)
Inclusion
Two study events with temporal relationship <ul style="list-style-type: none"> • (increase_decrease) (...*) (prepositions) (...*) • (...*) ((increase_decrease) (prepositions) and with) (...*)

• increase / decrease: increase, reduce, decrease, incidence
 • association words: associate, association, correlation, relationship
 • prepositions: in, among, after, following, of, for, on, to
 • ...*: all words except for words with special meanings

Identifying the study events

The phrases of “any words” captured from the last step would be mapped to UMLS Metathesaurus by MetaMap. The R code for extraction of MetaMap output was referred from Shah et al [15]. When the semantic types of the concepts in the MetaMap output are “Disease or Syndrome”, “Therapeutic or Preventive Procedure”, “Pharmacologic Substance” or other disease, procedure or medication related semantic types (priority were shown in Table 2), the concepts would be identified as study events. If there are more than two phrases matching to the specific semantic types above, the system would choose the first item of the output as the true study event.

Table 2 - Priority of event semantic types

Priority	Example semantic types
1	Disease or Syndrome Sign or Symptom Laboratory Procedure Therapeutic or Preventive Procedure Organic Chemical Pharmacologic Substance
2	Cell Function Organism Function Physiologic Function
3	Finding Qualitative Concept Quantitative Concept

Detecting the age and gender restriction criteria

Each sentence of eligibility criteria would be matched with 5 regular expressions (Table 4) for age restriction detection. In the example article, the sentence “*We excluded patients under the age of 20 years and those with a catastrophic illness certificate for autoimmune diseases other than Sjogren’s syndrome*” corresponded to the expression formula “(Comparing words) + the age of + (Value)” with the “excluded” concept, thus the age restriction is 20 years old for lower bound with no upper bound.

In addition, diabetes mellitus is one of the most common study events of NHIRD research, while the possible abbreviations of type 2 diabetes mellitus might be “T2DM”, “Type 2 DM”, “Type II DM” etc., which could not be detected because of more than one word in the parentheses. We will process the abbreviations of diabetes mellitus specially in the next step.

Table 4 - Regular expressions of age restriction criteria

- **((comparison_words) | (comparison_or_equal_to)) (value) (years | (...*) (and (comparison_words)) | (or (comparison_or_equal_to) (value) years))**
 - **between (value) and (value) years**
 - **(value) - (value) years**
 - **(value) years (...*) (and | or) (comparison_words)**
 - **((comparison_words) | (comparison_or_equal_to)) (age | (the age of)) (value)**
- **comparison words:** >, <, older than, elder than, more than, greater than, younger than, less than, over, above, under, below
 - **comparison or equal to:** >=, <=, older than or equal to, elder than or equal to, more than or equal to, greater than or equal to, younger than or equal to, less than or equal to, over or equal to, above or equal to, under or equal to, below or equal to, at least
 - **value:** [0-100]

The eligibility criteria are often written in the {Materials and Methods} section, thus before further analysis, the abbreviations in this section will be replaced with the corresponding full description, and the full width symbols will be replaced with halfwidth symbols. Other replacement rules are listed in Table 3. For the ease of analysis, we excluded the sentences which contain the phrases like “National Health Insurance Research Database” or “Data Availability”, and removed the descriptions of statistic analysis. The remain sentences would be regarded as eligibility criteria.

Each criteria sentence would be filtered with the keywords “ICD” and “code”, then the diagnosis code would be extracted with the regular expressions for ICD-9-CM code (Table 5). If the codes were written as a continuous code interval (e.g. “thyroid gland (240-242, 244-246)”), the start and end point would be annotated. The UMLS concept name and CUI of each code would be extracted by the R package ‘rUMLS’ (Table 6).

Table 5 - Regular expressions of ICD-9-CM code

- (((V | E | [0-9]) [0-9] [0-9]) (([.] (((X | x | [0-9]) (X | x | [0-9]) | *)) (*))))'

Table 6 - Continuity and UMLS CUI of ICD-9-CM code

Type	Symbols or term in the article	Replacement
Hyphen	- —	-
More than	≧ ≧ ≧	>=
Less than	≦ ≦ ≦	<=
Type 1 Diabetes Mellitus	T1DM Type 1 DM Type I DM	Type 1 Diabetes Mellitus
Type 2 Diabetes Mellitus	T2DM Type 2 DM Type II DM	Type 2 Diabetes Mellitus
Reference citation	Value within square brackets (e.g. [1], [1-2])	(Delete)

Code	Continuity	Concept name	CUI
240	Start	Simple and unspecified goiter	C0037158
242	End	Thyrototoxicosis with or without goiter	C0040156
244	Start	Acquired hypothyroidism	C0700502
246	End	Other disorders of thyroid	C0154164

Some papers would restrict the diagnosis records of study subjects made by specific specialists, in order to increase the accuracy and credibility. If there is the word “diagnosis” in the criteria sentence, the whole sentence will be further mapped to UMLS Metathesaurus by MetaMap. The words with the semantic type “Professional or Occupational Group” would be considered the restriction of specialists.

Results

Identifying the study type

In 280 papers, 99 belonged to the target study type “Increased or decreased (Event 2) risk after (Event 1)”. The 5-fold cross-validation results are depicted in Table 7. Papers like “Premotor Symptoms as Predictors of Outcome in Parkinson’s Disease: A Case-Control Study” [17] and “Irritable Brain Caused by Irritable Bowel? A Nationwide Analysis for Irritable Bowel Syndrome and Risk of Bipolar Disorder” [18] were misclassified due to matching to the exclusion regular expression.

Table 7 - System performance on study type identifying

Data set	Precision	Recall	F-measure
Training	0.772	0.818	0.794
Testing	0.778	0.818	0.789

Identifying the study events

In the 81 papers correctly classified as the target study type, the study events of 70 papers were correctly identified. 6 other papers were identified with only event 1 or event 2, and 5 papers failed. The event 1 of the article titled “A Longitudinal Study on Early Hospitalized Airway Infections and Subsequent Childhood Asthma” [19] was airway infections. However, it could not match to the UMLS Metathesaurus directly, thus it would be separated into “airway” and “infection” by MetaMap. Following the priority of event semantic type, the system would consider “infection” as event 1. After reviewing abstract, we found the real event 1 was acute bronchiolitis, that is to say, airway infections were a general description of their event 1.

There were 8 papers titled as “..... associated with”. However, the word “associate” did not imply temporal relationship between event 1 and event 2. For example, the order of events occurrence of “Rheumatoid Arthritis Risk Associated with Periodontitis Exposure: A Nationwide, Population-Based Cohort Study” [20] was (1) periodontitis (2) rheumatoid arthritis, while the order of “Hyperlipidemia Is Associated with Chronic Urticaria: A Population-Based Study” [21] was (1) hyperlipidemia (2) chronic urticaria. It goes to show that the temporal relationship between events should not be identified by article title only.

Furthermore, some phrases could not correctly map to UMLS Metathesaurus, leading to an error in events identifying. As shown in Table 8, the CUI of “Organophosphates Poisoning” was C0700359, however it would be separated into “Organophosphates” and “Poisoning”; “Urinary Stone Disease” would map to “Urologic Diseases”, “Calculi”, “Urinary Calculi” and “Disease”, although “Urinary Calculi” was closer to the real study event, “Urinary Calculi” was chosen because of its semantic type “Disease or Syndrome”.

Table 8 - Examples of incorrect events UMLS mapping

Input text	Mapping to UMLS Metathesaurus	Semantic type
Organophosphates Poisoning	Organophosphates	Organic Chemical
	Poisoning*	Injury or Poisoning
Urinary Stone Disease	Urologic Diseases*	Disease or Syndrome
	Calculi	Body Substance
	Urinary Calculi ^Δ	Body Substance
	Disease	Disease or Syndrome

^Δ: actual event

*: event chosen according to the priority of semantic type

Detecting the abbreviations

The system detected 545 abbreviations in the 99 target type papers, 525 of them were real abbreviations with corresponding full descriptions, 9 of them were correct abbreviations with wrong full descriptions, and finally, 11 papers were not abbreviations. Some papers would describe the comorbidities in sentence patterns like “chronic obstructive pulmonary disease (COPD; ICD-9-CM codes: 496)” to express abbreviations and operation codes at the same time. We have defined the abbreviations as only one word within the parentheses, thus these types of abbreviations would not be identified. On the contrary, some papers might describe the medication with only one operation code like “Antidepressants (N06A)”, which would lead “N06A” to be incorrectly seen as an abbreviation.

In addition, there were some abbreviations not created by the initials only, but also by letters inside the words, like “socioeconomic status (SES)”. When computing the similarity between real abbreviations and created abbreviations, based on the Jaro-Winkler distance, the similarity was 0.8 between “SES” and “SS” (from “socioeconomic status”), while 0.65 between “SES” and “S” (from “status”), hence “status” would be considered the full description of “SES”.

Detecting the age and gender restriction criteria

The 5-fold cross-validation results of age restriction detection of the 99 target papers are shown in Table 9. The most common criteria was “≥ 20 years old” (n = 23), and the second was “≥ 18 years old” (n = 18). We used the keywords “match” and “urbanization” to avoid detecting to the case matching description like “..... matched by the age (<20, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, >80 years)” or with the criteria of level of urbanization like “percentage of residents >65 years old”. In addition, one paper with criteria “younger than 36 months” was wrongly identified due to the system not converting the time unit in the preprocessing step.

In the development phase, the system detected the keywords of gender in eligibility criteria paragraph (Table 10). There were only 16 papers having the gender restriction, thus the false negative results would decrease the recall drastically. In addition, there were three papers titled with keywords about male or female reproductive system diseases thus the authors did not emphasize the gender restrictions and the system did not detect the restrictions in the paragraph.

Table 9 - System performance on age restriction detecting

Data set	Precision	Recall	F-measure
Training	0.968	0.952	0.960
Testing	0.969	0.950	0.959

Table 10 - System performance on gender restriction detecting

Data set	Precision	Recall	F-measure
Training	0.858	0.545	0.665
Testing	0.700	0.533	0.593

Extracting the ICD-9-CM codes

In the 99 target papers, there were 1575 ICD-9-CM diagnosis codes. 1438 of them were correctly extracted, the 144 lost were most due to (1) lack of “ICD” and “code” in the sentences, (2) written in ways like “401-5.x” which did not

meet our regular expression of ICD-9-CM codes, (3) the codes were written in tables rather than text content, (4) the continuous code were extracted but their start and endpoint were incorrect thus been considered as single codes. In addition, 67 items were mistaken for ICD-9-CM codes while they were actually the number of subjects.

Detecting the specialist restriction

The 5-fold cross-validation results of specialist restriction detection are shown in Table 11. The terms like “attending physician” were incorrectly considered as restrictions, thus decreased the precision.

Table 11 - System performance on specialist restriction

Data set	Precision	Recall	F-measure
Training	0.767	0.959	0.851
Testing	0.758	0.967	0.841

Conclusions

In this study, we focused on developing a pattern recognition-based method to extract the narrative eligibility criteria in NHIRD papers as simple as possible, which could identify the study type, medical concept, abbreviations, basic demographic characteristics and ICD-9-CM code restrictions in titles and contents of NHIRD articles. In this preliminary stage, we only used the articles from PLOS ONE for the ease of processing with XML files. The training and testing data were all from PLOS ONE at present. Moreover, we only dealt with the study type “Increased or decreased (Event 2) risk after (Event 1)” because the system could not identify the event relationship within article titles which include more than two events. In future work, we will include the 3,000 NHIRD papers across several journals as study materials, and establish the function of observation period and temporal relationship identifying, then convert the narrative text of scientific research articles into structured documents.

Acknowledgements

This work was supported by the Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data. Academia Sinica, Taipei, Taiwan.

References

- [1] R. Chin and B.Y. Lee, *Principles and practice of clinical trial medicine*, Elsevier, 2008.
- [2] Y.C. Chen, Y.C. Su, C.Y. Li, and S.K. Hung, 13-year nationwide cohort study of chronic kidney disease risk among treatment-naïve patients with chronic hepatitis B in Taiwan, *BMC Nephrol*, 2015; 16:110.
- [3] M.L. Pan, L.R. Chen, H.M. Tsao, and K.H. Chen, Relationship between Polycystic Ovarian Syndrome and Subsequent Gestational Diabetes Mellitus: A Nationwide Population-Based Study, *PLoS One*, 2015; 10:e0140544.
- [4] M.C. Lin, M.S. Lai, S.L. Jan, and Y.C. Fu, Epidemiologic features of Kawasaki disease in acute stages in Taiwan, 1997-2010: effect of different case definitions in claims data analysis, *J Chin Med Assoc*, 2015; 78:121-126.
- [5] M.J. Chang, H.I. Ma, and T.H. Lu, Estimating the prevalence of cerebral palsy in Taiwan: A comparison of different case definitions, *Res Dev Disabil*, 2014; 36C: 207-212.
- [6] W.L. Lee, W.H. Chang, K.C. Wang, C.Y. Guo, Y.J. Chou, N. Huang, H.Y. Huang, M.S. Yen, and P.H. Wang, The Risk of Epithelial Ovarian Cancer of Women With Endometriosis May be Varied Greatly if

Diagnostic Criteria Are Different: A Nationwide Population-Based Cohort Study, *Medicine (Baltimore)*, 2015; 94:e1633.

- [7] C.L. Meinert, *ClinicalTrials: Design, Conduct and Analysis*, Oxford University Press, 2012.
- [8] Y. Ni, J. Wright, J. Perentesis, T. Lingren, L. Deleger, M. Kaiser, I. Kohane, and I. Solti, Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients, *BMC Med Inform Decis Mak*, 2015; 15:28.
- [9] K. Milian, A. Bucur, and A.t. Teije, Formalization of clinical trial eligibility criteria-Evaluation of a pattern-based approach, in: *Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference*, 2012; 1-4.
- [10] Z. He, S. Carini, T. Hao, I. Sim, and C. Weng, A method for analyzing commonalities in clinical trial target populations, *AMIA Annu Symp Proc*, 2014; 1777-1786.
- [11] W.W. Chapman, D. Chu, and J.N. Dowling, ConText: An Algorithm for Identifying Contextual Features from Clinical Text, *Prague, Czech Republic*, 2007; 81-88.
- [12] R.J. Byrd, S.R. Steinhubl, J. Sun, S. Ebadollahi, and W.F. Stewart, Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records, *Int J Med Inform*, 2014; 83:983-992.
- [13] M.C. Lu, W.Y. Yin, T.Y. Tsai, M. Koo, and N.S. Lai, Increased risk of primary Sjogren's syndrome in female patients with thyroid disorders: a longitudinal population-based study in Taiwan, *PLOS ONE*, 2013; 8:e77210.
- [14] A.R. Aronson and F.M. Lang, An overview of MetaMap: historical perspective and recent advances, *J Am Med Inform Assoc*, 2010; 17:229-236.
- [15] N.H. Shah, N. Bhatia, C. Jonquet, D. Rubin, A.P. Chiang, and M.A. Musen, Comparison of concept recognizers for building the Open Biomedical Annotator, *BMC Bioinformatics*, 2009; 10(Suppl 9):S14.
- [16] A.S. Schwartz and M.A. Hearst, A simple algorithm for identifying abbreviation definitions in biomedical text, *Pac Symp Biocomput*, 2003; 451-462.
- [17] Y.H. Wu, W.J. Lee, Y.H. Chen, M.H. Chang, and C.H. Lin, Premotor Symptoms as Predictors of Outcome in Parkinsons Disease: A Case-Control Study, *PLOS ONE*, 2016; 11:e0161271.
- [18] C.J. Liu, L.Y. Hu, C.M. Yeh, Y.W. Hu, P.M. Chen, T.J. Chen, and T. Lu, Irritable brain caused by irritable bowel? A nationwide analysis for irritable bowel syndrome and risk of bipolar disorder, *PLOS ONE*, 2014; 10:e0121906.
- [19] M.J. Jeng, Y.S. Lee, P.C. Tsao, C.F. Yang, and W.J. Soong, A longitudinal study on early hospitalized airway infections and subsequent childhood asthma, *PLoS One* 10 (2014), e0121906.
- [20] Y.Y. Chou, K.L. Lai, D.Y. Chen, C.H. Lin, and H.H. Chen, Rheumatoid Arthritis Risk Associated with Periodontitis Exposure: A Nationwide, Population-Based Cohort Study, *PLOS ONE*, 2015; 10:e0139693.
- [21] S.D. Chung, K.H. Wang, M.C. Tsai, H.C. Lin, and C.H. Chen, Hyperlipidemia Is Associated with Chronic Urticaria: A Population-Based Study, *PLoS One*, 2016; 11:e0150304.

Address for correspondence

Mei-Lien Pan: mlpan66@gmail.com

Assessing the Representation of Occupation Information in Free-Text Clinical Documents Across Multiple Sources

Elizabeth A. Lindemann^a, Elizabeth S. Chen^b, Sripriya Rajamani^{c, d},
Nivedha Manohar^e, Yan Wang^c, Genevieve B. Melton^{a, c}

^a Department of Surgery, University of Minnesota, Minneapolis, MN, USA,

^b Center for Biomedical Informatics, Brown University, Providence, RI, USA,

^c Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA,

^d Public Health Informatics Program, University of Minnesota, Minneapolis, MN, USA,

^e West Side Community Health Services, St. Paul, MN, USA

Abstract

There has been increasing recognition of the key role of social determinants like occupation on health. Given the relatively poor understanding of occupation information in electronic health records (EHRs), we sought to characterize occupation information within free-text clinical document sources. From six distinct clinical sources, 868 total occupation-related sentences were identified for the study corpus. Building off approaches from previous studies, refined annotation guidelines were created using the National Institute for Occupational Safety and Health Occupational Data for Health data model with elements added to increase granularity. Our corpus generated 2,005 total annotations representing 39 of 41 entity types from the enhanced data model. Highest frequency entities were: Occupation Description (17.7%); Employment Status – Not Specified (12.5%); Employer Name (11.0%); Subject (9.8%); Industry Description (6.2%). Our findings support the value of standardizing entry of EHR occupation information to improve data quality for improved patient care and secondary uses of this information.

Keywords:

Occupations; Social Determinants of Health; Electronic Health Records.

Introduction

The need to accurately capture occupation and other social history information is central for the use of this information in the provision of direct clinical care and for secondary uses like research and risk stratification of patients [1]. As electronic health record (EHR) system use increases within healthcare organizations, there is an opportunity for improved capture of this information at the point of care. Within the United States, the National Academy of Medicine (NAM) in their 2011 report “Incorporating Occupational Information in Electronic Health Records” advocates for inclusion of social factors in EHR systems, due to the important impact of social factors on health status and outcomes [2]. Similarly, Phases 1 and 2 of their 2014 reports “Capturing Social and Behavioral Domains and Measures in Electronic Health Records” [3, 4] further emphasize the importance of these factors. The National Institute for Occupational Safety and Health (NIOSH), a leader in efforts to qualify occupation information in the United States, has done extensive work to promote documentation of occupation information in a standard manner by creating the Occupational Data for Health (ODH) data model [5].

Prior work has been done to broadly analyze how social history information is captured in EHR systems [6, 7], public health datasets [8], and how occupation-related information appears within standards and within social history in the EHR [9, 11]. More specifically, using a top-down approach, Rajamani *et al.* examined reports, standards, surveys, and research measures to analyze the ODH model’s ability to provide coverage for occupation-related information [9], including the Health Level Seven Clinical Document Architecture content module of the Integrating the Healthcare Enterprise Patient Care Coordination Technical Framework, which incorporates the ODH model [10]. Aldekhyyel *et al.* examined the content and quality of entries within the free-text occupation field of the Fairview Health Services enterprise EHR social history module [11]. The results of these studies provided a comprehensive overview of the current state of representation and standardization of occupation information. Collectively, they highlight that the ODH model is robust in coverage of occupation-related information and that the content and quality of occupation information within the EHR can be inconsistent and variable.

The main objective of this study was to build upon prior efforts of this group of authors by looking at a range of free-text clinical sources to further inform occupation representations leveraging the NIOSH ODH model. In performing our analysis, we anticipate additional refinements thus enabling future standardization and additional insights into language around occupation. These will ultimately aid in standards refinement and future natural language processing (NLP) efforts around occupation information.

Methods

Data Sources

This study utilized six clinical document sources comprised of a mix of both publicly available and local sources to analyze free-text mentions of occupation and related information within notes. Information was unstructured, yielding a large variety of occupation-related information. The public domain note sources used came from 491 “Consult – History and Physical” notes from MT Samples (MTS) [12] and 200 de-identified “H&P” (History and Physical Examination) notes from the University of Pittsburgh Medical Center (UPMC) [13] obtained via a data use agreement. Four document types were analyzed from the University of Minnesota (UMN)-affiliated Fairview Health Services Epic EHR: (1) Social History Documentation, (2) So-

cial Work notes, (3) Physical Therapy notes, and (4) Occupational Therapy notes available through the University of Minnesota Clinical Data Repository (CDR) from 2013.

Analysis of Clinical Text

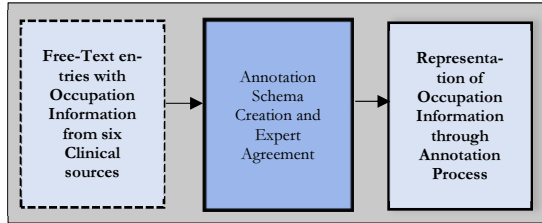


Figure 1 – Overview of Assessment of Occupation Information in Clinical Text

Clinical text analysis consisted of three main parts: (1) text deidentification, (2) schema creation, and (3) text annotation. Sentences from Fairview Health Services were anonymized prior to annotation using the Safe Harbor method [14]. Codes for anonymizing data came from previous work [15]. Further obfuscation was used where necessary to avoid personally identifiable data, and descriptors were created to provide specificity regarding employer type. These included descriptors for level of government employers, hospitality industry employers, healthcare employers, education employers, and Fortune 500 employers. Fortune 500 companies were determined using the Fortune 500 yearly ranking for 2016 [16]. Any remaining companies that did not fit into more specific descriptors were anonymized to “company.”

Annotations were made using the brat rapid annotation tool (BRAT) [17]. Annotation schema and guidelines were derived from the NIOSH ODH model categories and elements, providing the parent entities *Occupational History*, *Usual Occupation*, *Employment Status*, *Occupational Injury*, and *Occupational Exposure* and associated child entities. The Systemized Nomenclature of Medicine-Clinical Terms (SNOMED CT) [18] was examined, but it was determined that the ODH model was more relevant for our corpora, given its usage in previous work by this group of authors [9]. An initial set of 25 sentences was annotated with a group of annotators, and 50 overlapping sentences were individually annotated to evaluate the

annotation schema and guidelines (Figure 1) by three annotators (EL, SR, NM). Additional entities were created to increase coverage, including *Subject*, *Negation*, *Temporal*, *Occupational Conditions*, and *Occupation Status*. The final schema had 9 parent entities and 32 child entities, for a total of 41 elements (Table 1).

Annotators were instructed to annotate at the most specific level of detail. For example, in the sentence, “She is a registered nurse,” the text “registered nurse” is annotated as the child entity *Occupation Description*, rather than the parent entity *Occupational History*. Parent entities were used as general annotation categories when a child entity could not be specified. In another example, “Prior to retirement, pt worked as a civil engineer,” the sentence contains annotations relating to both clauses (i.e., “Prior to” is *Temporal – Time Frame*; “retirement” is *Employment Status – Retired*; “worked” is *Employment Status – Not Specified*; “civil engineer” is *Occupational Description*). Relationships were created to describe how entities are connected to each other. An overlapping set of 10% of sentences was annotated to calculate inter-rater reliability between three annotators (achieving a Cohen’s kappa of 0.76 and proportion agreement of 0.95). When inter-rater reliability was ascertained, the remaining sentences were annotated using the most expanded version of the schema. Following annotation process, annotations were extracted by element, generating a list of values for each element. These values were grouped by similar meaning for highest frequency elements. For example, for *Occupation Description*, the group ‘Legal Occupations’ represents annotations ‘paralegal,’ ‘attorney,’ and ‘tax attorney.’

Results

A total of 868 sentences from the six sources of clinical documents were annotated, yielding 2,005 annotations, which were mapped to 41 entities. The most frequent entities were: *Occupation Description* (17.7%), *Employment Status – Not Specified* (12.5%), *Employer Name* (11.0%), *Subject* (9.8%), and *Industry Description* (6.2%). Table 1 summarizes the representation of entities across source type. The Fairview Social History Documentation sentences have the greatest variety of occupation information, containing 39 (95%) of 41 possible entities. The sentences from the Fairview Physical Therapy Notes have the least diversity in occupation-related information, containing annotations for 13 (32%) of the 41 possible entities. No annotations were made for the entities *Volunteer* and *Usual Occupation*.

Table 1 – Distribution of occupation entities across sources. ([# of sentences per source]; *n=number of annotations; SH=Social History; SW=Social Work; OT=Occupational Therapy; PT=Physical Therapy)

Entities	Fairview SH Documentation						All Sources [868] (n=2005)*
	MTS [171] (n=385)*	UPMC [44] (n=101)*	UPMC [553] (n=1295)*	Fairview SW Notes [58] (n=113)*	Fairview OT Notes [16] (n=65)*	Fairview PT Notes [26] (n=46)*	
Occupational History	-	-	2 (0.2%)	-	1 (1.5%)	-	3 (0.1%)
Industry Description	22 (5.7%)	10 (9.9%)	80 (6.2%)	6 (5.3%)	5 (7.7%)	2 (4.3%)	125 (6.2%)
Occupation Description	77 (20.0%)	24 (23.8%)	227 (17.5%)	15 (13.3%)	3 (4.6%)	8 (17.4%)	354 (17.7%)
Job Duties	13 (3.4%)	-	55 (4.2%)	3 (2.7%)	3 (4.6%)	2 (4.3%)	76 (3.8%)
Employer Name	36 (9.4%)	7 (6.9%)	151 (11.7%)	18 (16.0%)	1 (1.5%)	8 (17.4%)	221 (11.0%)
Employer Location	7 (1.8%)	4 (4.0%)	48 (3.7%)	3 (2.7%)	-	3 (2.2%)	65 (3.2%)
Usual Occupation	-	-	-	-	-	-	-
Usual Industry Description	1 (0.3%)	-	3 (0.2%)	-	-	-	4 (0.2%)
Usual Occupation Description	5 (1.3%)	1 (1.0%)	4 (0.3%)	-	-	-	10 (0.5%)
Occupation Status	1 (0.3%)	-	3 (0.2%)	-	-	-	4 (0.2%)
Employment Status	4 (1.0%)	-	1 (0.1%)	-	-	-	5 (0.2%)
Full-Time	2 (0.5%)	1 (1.0%)	24 (1.9%)	4 (3.5%)	-	6 (13.0%)	37 (1.8%)
Part-Time	3 (0.8%)	-	9 (0.7%)	4 (3.5%)	-	1 (2.2%)	17 (0.8%)
Not Specified	46 (11.9%)	13 (12.9%)	141 (10.9%)	4 (3.5%)	41 (63.1%)	6 (13.0%)	251 (12.5%)
Self-Employed	4 (1.0%)	1 (1.0%)	2 (0.2%)	2 (1.8%)	1 (1.5%)	1 (2.2%)	11 (0.5%)
Employed but temporarily not working	1 (0.3%)	1 (1.0%)	1 (0.1%)	4 (3.5%)	1 (1.5%)	-	8 (0.4%)
Not Employed	12 (3.1%)	10 (9.9%)	30 (2.3%)	3 (2.7%)	-	-	55 (2.7%)

Retired	29 (7.5%)	-	56 (4.3%)	1 (0.9%)	-	-	86 (4.3%)
Disability (Not employed)	9 (2.3%)	-	4 (0.3%)	1 (0.9%)	-	-	14 (0.7%)
Military Service	2 (0.5%)	1 (1.0%)	1 (0.1%)	-	-	-	4 (0.2%)
Student	-	-	22 (1.7%)	-	-	-	22 (1.1%)
Student Type	1 (0.3%)	-	19 (1.5%)	-	-	-	20 (1.0%)
Student Status	5 (1.3%)	-	9 (0.7%)	1 (0.9%)	-	-	15 (0.7%)
Student Other	3 (0.8%)	-	55 (4.2%)	1 (0.9%)	1 (1.5%)	2 (4.3%)	62 (3.1%)
Volunteer	-	-	-	-	-	-	-
Not in Paid Workforce (Child)	3 (0.8%)	-	25 (1.9%)	-	-	-	28 (1.4%)
Homemaker/Housewife	8 (2.1%)	-	30 (2.3%)	1 (0.9%)	1 (1.5%)	-	40 (2.0%)
Caretaker	2 (0.5%)	-	6 (0.5%)	-	-	-	8 (0.4%)
Occupational Injury	2 (0.5%)	1 (1.0%)	3 (0.2%)	-	-	-	6 (0.3%)
Occupational Exposure	1 (0.3%)	2 (2.0%)	2 (0.2%)	-	-	-	5 (0.2%)
Occupational Conditions	8 (2.1%)	-	7 (0.5%)	3 (0.2%)	1 (1.5%)	4 (8.7%)	23 (1.1%)
Temporal	1 (0.3%)	-	3 (0.2%)	-	-	-	4 (0.2%)
Work Schedule	3 (0.8%)	-	5 (0.4%)	2 (1.8%)	1 (1.5%)	2 (4.3%)	13 (0.6%)
Start Date	4 (1.0%)	-	18 (1.4%)	2 (1.8%)	-	-	24 (1.2%)
End Date	2 (0.5%)	-	3 (0.2%)	-	-	-	5 (0.2%)
Days Per Week	-	1 (1.0%)	5 (0.4%)	-	1 (1.5%)	-	7 (0.3%)
Hours Per Week	-	-	3 (0.2%)	-	1 (1.5%)	1 (2.2%)	5 (0.2%)
Duration Years	1 (0.3%)	3 (3.0%)	10 (0.8%)	5 (4.4%)	1 (1.5%)	-	33 (1.6%)
Time Frame	38 (9.9%)	20 (19.8%)	55 (4.2%)	6 (5.3%)	1 (1.5%)	-	120 (6.0%)
Subject	10 (2.6%)	1 (1.0%)	162 (12.5%)	23 (20.4%)	1 (1.5%)	-	197 (9.8%)
Negation	6 (1.6%)	-	11 (0.8%)	1 (0.9%)	-	-	18 (0.9%)
# of Entities	35	16	39	23	17	13	39

Tables 2 through 5 provide summaries of value sets for the most commonly represented elements, demonstrating how elements were represented in clinical text. Table 2 summarizes the value sets for the most common entity, *Occupation Description*. Values are grouped according to the Bureau of Labor Statistics 2010 Standard Occupation Classification major groups [19]. Of the major groups in SOC, 22 of 23 are represented across

sources. Military Specific Occupations is the only group not represented within these clinical sources, due to its small presence within the corpus, for *Occupation Description* annotations. Healthcare Practitioners and Technical Occupations are most common with 73 overall values (20.6%) and 51 unique values. Farming, Fishing, and Forestry Occupations are the least commonly represented, with 1 entry (0.3%).

Table 2 – Distribution of values for *Occupation Description* element with grouping based on the 2010 Standard Occupation Classification groups [16]

2010 SOC Classification Group (n=23)	Occupation Description		
	Number of Total Values (n=354)	Number of Unique Values (n=267)	Frequency (%)
Healthcare Practitioners and Technical Occupations	73	51	20.6
Business and Financial Operations Occupations	40	39	11.3
Education, Training, and Library Occupations	36	31	10.2
Construction	22	18	6.2
Management Occupations	21	15	5.9
Food Preparation and Serving Related Occupations	20	11	5.6
Office and Administrative Support Occupations	19	16	5.4
Architecture and Engineering Occupations	17	7	4.8
Installation, Maintenance, and Repair Occupations	15	11	4.2
Sales and Related Occupations	11	10	3.1
Transportation and Material Moving Occupations	11	2	3.1
Building and Grounds Cleaning and Maintenance Occupations	9	5	2.5
Personal Care and Service Occupations	8	6	2.3
Community and Social Service Occupations	8	6	2.3
Computer and Mathematical Occupations	7	7	2.0
Arts, Design, Entertainment, Sports, and Media Occupations	7	6	2.0
Legal Occupations	7	3	2.0
Life, Physical, and Social Science Occupations	6	6	1.7
Protective Service Occupations	6	6	1.7
Healthcare Support Occupations	5	5	1.4
Production Occupations	4	4	1.1
Farming, Fishing, and Forestry Occupations	1	1	0.3
Unclassified	1	1	0.3

The entity *Employer Status – Not Specified* is a child entity of *Occupation Status* that is heavily used throughout the annotation process, because annotation guidelines specified to annotate all tenses of the verb “work” as *Employment Status – Not Specified*. This was done to capture work behaviors when part-time or full-time employment status is not directly stated (Table 3). Consequently, “Work” is the most frequent group in this

data set, with 214 total values (86.3%) and only 6 unique values. The “Other” group included a variety of other terms.

Table 3 – Distribution of values for Employment Status – Not Specified element

Group (n=2)	Employment Status – Not Specified		Frequency (%)
	Number of Total Values (n=248)	Number of Unique Values (n=13)	
Work	214	6	86.3
Other	34	7	13.7

indicating work status, including “released to regular work” and “has been on light duty.

Table 4 summarizes the value sets for the entity *Employer Name*. All values from Fairview Health Services Epic EHR in this set were anonymized prior to annotation and are grouped accordingly. Educational Services was the most frequently occurring group with 83 entries (37.6%). Student status is a contributing factor to this result, as schools were annotated as *Employer Name*. Military employment is least commonly seen with 1 entry (0.5%). Five entries (2.3%) did not provide enough information in order to group by employer type.

Table 4 – Distribution of values for Employer Name element

Group (n=7)	Employer Name		Frequency (%)
	Number of Total Values (n=221)	Number of Unique Values (n=13)	
Educational Services	83	6	37.6
Private and publicly held companies that do not qualify for Fortune 500 ranking	47	7	21.3
Health Care and Social Assistance	30	6	13.6
Fortune 500 Companies	25	5	11.3
Public Administration	19	4	8.6
Accommodation and Food Services	11	3	5.0
Not enough information for classification	5	2	2.3
Military	1	1	0.5

Table 5 summarizes the value sets for the entity *Subject*. This entity was added to the annotation schema after initial evaluation when a need for granularity was discovered. Parent group entries are most frequent with 137 total entries (74.5%) and 13 unique entries. Names of individuals and references to spouses were also common – de-identified prior to annotation as “Name” – 15.2% and 7.1% respectively.

Table 5 – Distribution of values for Subject element

Group (n=4)	Subject		Frequency (%)
	# of Total Values (n=184)	# of Unique Values (n=28)	
Parent	137	13	74.5
Name	28	3	15.2
Spouse	13	6	7.1
Other Subject	6	6	3.3

Discussion

This study’s findings demonstrate the wide variety and range of occupation-related information within EHR clinical texts. A total of 2,005 annotations were made for occupation-related information in 868 sentences from six clinical document sources. This pervasiveness of occupation-related information within

free-text notes points to the need for standardized entry within the EHR. In large part, the NIOSH ODH model proves appropriate for representing most of the occupation information, but several additions to the schema were necessary for comprehensive representation in our clinical text corpus. Specifically, the *Subject* entity is one of the most frequently occurring entities (9.8%) and was added during initial schema evaluation when need was assessed. Within this element, parent group entries are highly used, demonstrating a clinical interest in parental occupation and a prevalence of pediatric patients in this dataset. This correlation could also point to the relevance of a parent’s occupation on a child’s health outcomes. Similarly, the prevalence and interest in spousal occupations within this dataset points to the relevance of a spouse’s occupation to an individual’s health outcomes.

The Fairview Social History Documentation notes represent the largest portion of the dataset, with 553 sentences and 1,295 annotations. This source also presents the greatest coverage of occupation-related elements, with 39 of 41 possible elements across the dataset. *Volunteer* and *Usual Occupation* (parent entity) are the two elements not represented in the dataset; these elements were also not seen in any other source. The Fairview Physical Therapy Notes present the least coverage of elements with 13 of 41 possible elements. The Fairview Occupational Therapy Notes and Fairview Social Work Notes also present a more focused coverage of elements with 17 and 23 elements, respectively. This could be due to the specialty nature of these notes and more focused reasons for obtaining occupation information. The MTS sentences presented the greatest coverage of elements per sentence annotated with 35 elements represented in 171 sentences.

The addition of the element *Occupational Conditions* was useful in describing daily stressors of an individual’s job that have noted long-term effects on health outcomes, but do not necessarily constitute injury or exposure. Among these are items such as “Pt describes standing 12 hour days on concrete for his job,” “Patient works at a computer and on the phone all day,” and “He is employed in sales, which requires quite a bit of walking, but he is not doing any lifting.” While *Occupational Conditions* annotations comprise only 1.1% of total annotations, there were zero duplicate entries, indicating the scope of conditions – both physical and mental demands – individuals face in their occupations.

Occupation Description is the most frequent element seen in this dataset with 354 total values in 868 sentences. Among these values, 267 were unique and all but one were categorized into 22 of 23 SOC major groups. The ability to classify a large number of unique entries into 22 groups proves that occupation descriptions could benefit from standardized entry of this information within the EHR. While value sets are wide reaching, many occupations fall into a smaller set of groups. *Employer Name* was also pervasive in the dataset with 221 total values. These were categorized into 7 groups based on anonymized codes.

We also observed that ambiguity was common within the *Occupation Status* related entities, such as *Self-Employed* and *Homemaker/Housewife*. For example, some individuals may hold more than one type of employment status (i.e., both a full-time position and a part-time position). Some individuals may have overlapping employment status (i.e., the term “stay at home mom” implies both a caretaker and homemaker role). Self-employment also presents ambiguity, as this could refer to either full-time or part-time employment status in addition to self-employment.

The fifth most frequent element within the dataset was *Industry Description* (6.2%). This value set was not grouped by any classification set because a large majority of entries were unique and presented varying degrees of information presented. The issue of discrepancies in information is best represented by *Industry Description* but was prevalent throughout the dataset. This underscores the lack of standardization currently in place in documenting occupation information within the EHR and the need for further work to characterize the wealth and variety of occupation information that has potential impacts on health status and outcomes.

This work builds off of previous work [8, 9] to identify how occupation information is represented across the literature and various aspects of the EHR. Work could be done involving comparisons against occupation information in EHR systems from other vendors. Future work will focus on dissemination of this research on the Brown Digital Repository [20], creating granularity within the current model, and lead to better NLP techniques to analyze occupation information. This will contribute towards standardized entry of occupation information within the EHR, promoting data quality, and ultimately improving patient care and secondary use of occupation information.

Conclusion

As EHR system use becomes more widespread it will become imperative to have standardized entry of factors that influence health status and outcomes. Several respected groups have recognized occupation as a factor of health status. With this recognition comes the need to understand and standardize how occupation-related information is being captured within the EHR. This study analyzed free-text clinical notes from a variety of sources in order to characterize the state of occupation-related information within the EHR. The NIOSH ODH data model proved robust in characterizing information content, and additions were made to the annotation schema to provide additional granularity. This work has potential to lead to more detailed and knowledgeable standards and provides a basis for creating a standardized entry system within the EHR and improved NLP techniques.

Acknowledgements

This work was supported in part by National Library of Medicine grant R01LM011364.

References

- [1] H.L. Behfourouz, P.K. Drain, J.J. Rhatigan, Rethinking the Social History, *N Engl J Med*, **371**:14 (2014), 1277-1279.
- [2] National Academy of Medicine (NAM), Incorporating Occupational Information in Electronic Health Records: Letter Report [Internet]. Washington, DC, 2011. [cited 2016 December 9] Available from: http://www.nap.edu/catalog.php?record_id=13207.
- [3] National Academy of Medicine (NAM), Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1. 2014. [cited 2016 December 9]. Available from: <http://www.nationalacademies.org/hmd/Reports/2014/Capturing-Social-and-Behavioral-Domains-in-Electronic-Health-Records-Phase-1.aspx>.
- [4] National Academy of Medicine (NAM), Capturing Social and Behavioral Domains in Electronic Health Records: Phase 2. 2014. [cited 2016 December 9]. Available from: <http://www.nationalacademies.org/hmd/Reports/2014/EHRdomains2.aspx>.
- [5] National Institute for Occupational Safety and Health (NIOSH), Structuring Patient Work Information in EHRs to Improve Patient Care and Public Health; Occupational Data for Health (ODH) Model, Public Health Informatics Conference, Atlanta, GA, 2014.
- [6] E.S. Chen, S. Manaktala, I.N. Sarkar, G.B. Melton, A multi-site content analysis of social history information in clinical notes, *AMIA Annu Symp Proc*, (2011), 227-36.
- [7] T.J. Winden, E.S. Chen, E. Lindemann, Y. Wang, E.W. Carter, G.B. Melton, Evaluating living situation, occupation, and hobby/activity information in the electronic health record, *AMIA Annu Symp Proc*, (2014), 139.
- [8] G.B. Melton, S. Manaktala, I.N. Sarkar, E.S. Chen, Social and Behavioral History Information in Public Health Datasets, *AMIA Annu Symp Proc*, (2012), 625-634.
- [9] S. Rajamani, E.S. Chen, R. Aldekhayel, Y. Wang, G.B. Melton, Validating the Occupational Data for Health Model: An Analysis of Occupational Information in Reports, Standards, Surveys, and Measures, *AMIA Annu Symp Proc*, (2016) 115-116.
- [10] Health Level Seven (HL7) – IHE PCC Technical Committee. Integrating the Healthcare Enterprise (IHE) Patient Care Coordination (PCC) Technical Framework Supplement – CDA Content Modules: Trial Implementation Guide [Internet]. [cited 2017 April 6]. Available from: http://www.ihe.net/uploadedFiles/Documents/PCC/IHE_PCC_TF_Vol1.pdf.
- [11] R. Aldekhayel, E.S. Chen, S. Rajamani, Y. Wang, G.B. Melton, Content and Quality of Free-Text Occupation Documentation in the Electronic Health Record, *AMIA Annu Symp Proc*, (2016), 1708-1716.
- [12] MTSamples, MTSamples.com [Internet]. [cited 2016 December 9]. Available from: <http://mtsamples.com>.
- [13] University of Pittsburgh NLP Repository. Department of Biomedical Informatics, University of Pittsburgh. [Internet]. [cited 2016 December 9]. Available from: <http://www.dbmi.pitt.edu/nlpfront>.
- [14] U.S. Department of Health and Human Services, Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule [Internet]. [cited 2016 December 9]. Available from: <http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>.
- [15] S. Moon, S. Pakhomov, G. Melton, Clinical Abbreviation Sense Inventory [Internet]. [cited 2016 December 9]. Available from: http://conservancy.umn.edu/bitstream/handle/11299/137704/README_DeidentifiedSymbolDataSet.txt?sequence=2&isAllowed=y.
- [16] Fortune 500, Fortune 500 List [Internet]. [cited 2016 December 11] Available from: <http://beta.fortune.com/fortune500/list>.
- [17] Brat: Brat rapid annotation tool [Internet]. [cited 2016 December 9]. Available from: <http://brat.nlplab.org>.
- [18] SNOMED International SNOMED CT Browser [Internet]. [cited 2017 April 6] Available from: <http://browser.ihtsdo.org/>.
- [19] Bureau of Labor Statistics, The 2010 Standard Occupational Classification System [Internet]. [cited 2016 December 9]. Available from: http://www.bls.gov/soc/major_groups.htm.
- [20] Brown University Libraries, SFHERE Project [Internet]. [cited 2017 April 7]. Available from: https://repository.library.brown.edu/studio/collections/id_700/.

Address for Correspondence:

Genevieve B. Melton, MD, PhD, FACMI
Department of Surgery and Institute for Health Informatics
University of Minnesota
MMC 450, 420 Delaware Street SE, Minneapolis, MN, 55455,
USA
Email: gmelton@umn.edu

Precision Cohort Finding with Outcome-Driven Similarity Analytics: A Case Study of Patients with Atrial Fibrillation

Haifeng Liu^a, Xiang Li^a, Guotong Xie^a, Xin Du^b, Ping Zhang^c,
Chengming Gu^d, Jingyi Hu^d

^aIBM Research China, Beijing, China

^bAnzhen Hospital, Beijing, China

^cIBM T. J. Watson Research Center, Yorktown Heights, NY, USA

^dPfizer China, Beijing, China

Abstract

Dividing patients into similar groups plays a significant role in implementing personalized care. Clinicians and researchers have been applying patient grouping techniques in disease phenotyping, risk stratification, and personalized medicine. However, the current approaches are either based on pure domain knowledge where the underlying patient similarity cannot be precisely quantified, or based on unsupervised clustering techniques which completely ignore the clinical context of measuring patient similarity. In the study, we propose an outcome-driven approach to identify clinically similar patients which are grouped together as a precision cohort. The approach quantitatively measures the similarity between patients in terms of a particular clinical outcome of interest, thus patients who have a similar clinical outcome tend to be grouped into the same group. We demonstrate the effectiveness of the approach in a real-world case study: from an atrial fibrillation patient cohort that is usually considered to be at high risk for ischemic stroke (IS), according to current clinical guidelines. Our approach successfully identified a precision cohort of patients with truly low risk of IS.

Keywords:

Cluster Analysis; Machine Learning; Precision Medicine

Introduction

In order to understand complex disease conditions and to provide personalized care, it is crucial for clinicians to divide patients into subgroups such that patients in one group are similar to each other. Successful patient grouping is particularly beneficial for the tasks of disease phenotyping, risk stratification and personalized medicine. In current clinical research and/or practice, patient grouping is generally conducted based on some scoring schemes recommended by clinical guidelines, which stratify patients into groups with different levels of risk. The patients with the same level of risk are considered to be similar. This mechanism may lead to some problems: 1) a general clinical guideline may not fit for the local clinical practice or population; 2) the similarity between patients with the same risk level are not quantified because the scoring schema is generated for a population and the detailed patient conditions are not differentiated in the same level.

In recent years, unsupervised clustering has been applied to identify groups of patients with different phenotypes and implement personalized medicine [1;2] in clinical research. Although the method can be adapted to local clinical practice and quantify the similarity between patients using distances between vectors of patient features, it ignores the fact that

patient similarities are usually context-based, i.e., the similarity degree between two patients' conditions may vary in terms of particular clinical outcomes of interest. For example, for three patients A, B and C with atrial fibrillation (AF), clinicians would regard that A and B are more similar in terms of stroke-occurrence risk, and A and C are more similar to each other when considering myocardial infarction (MI) risk. This is because the impacts of risk factors on these two outcomes are totally different (e.g., smoking and body mass index play important roles in the risk of MI while their impacts on stroke are relatively smaller).

This study aims to demonstrate how an outcome-driven similarity analytics method can be utilized to address the issues above. With the given clinical outcome of interest, the key idea is to cluster patients into groups based on a learned similarity (distance) metric from patients' clinical records where patients with the same clinical outcome are considered to be similar. One related work [3] has been reported using a learned distance metric to retrieve the K most similar patients and providing the prognosis insight based on the physiological time-series data of similar patients. Another work [4] is to learn the similarity metric from physicians' feedbacks and use patient similarity for decision support. Our work differs from them in using the learned distance metric to divide the patients into groups, followed by identifying the characteristics of similar patients in each group. To further understand the resultant groups, we also discover the discriminating rules between groups. The discovered rules can guide clinicians to easily assign new patients into their similar patient group, and we call such a patient group a precision cohort. Personalized care can then be recommended to the patients based on the insights discovered from this precision cohort. An alternative approach to divide patients into subgroups without similarity metric learning and clustering is to directly build a decision tree to split the patients against the outcome [5]. However, it suffers from two limitations: 1) similar patients could be dispersed in different branches, thereby requiring manual regrouping after the tree is built; 2) there is no distance metric for measuring the exact similarity between two patients.

We validate our approach in a real-world case study where we stratify a population of AF patients with high risk of ischemic stroke (IS) into a few groups and identify a particular group of patients with truly low risk. With the demonstrated effectiveness, we believe that other diseases and scenarios of finding precision cohorts could generally benefit from the proposed approach.

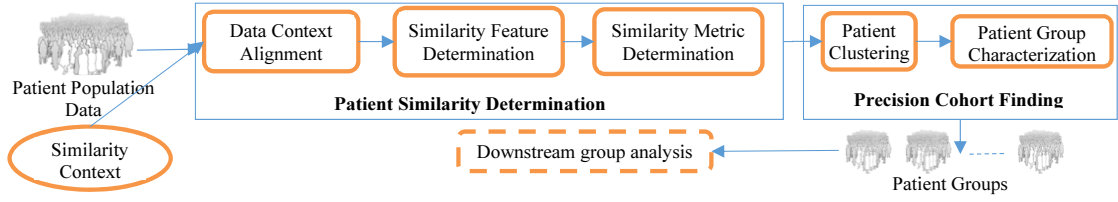


Figure 1 - Methodology of outcome-driven patient similarity analysis

Methods

As precise patient similarity is context-dependent, and varies along the patient conditions, outcomes of interest, and particular clinical scenarios, we first define patient similarity context as follows. A patient similarity context consists of:

- A target patient population of study, e.g., the patients with diagnosis of diabetes type 2 but without any other complications,
- A clinical outcome of interest, e.g., mortality in two years, re-hospitalization in six months, and so on. The assumption is that two patients with the same outcome are considered to be similar, and
- A clinical scenario when the similarity analytics need to be done, e.g., when patients are first diagnosed with type 2 diabetes, when patients are hospitalized, or when patients are registered into a particular study.

Figure 1 illustrates the overall methodology of how we find precision cohorts by outcome-driven patient similarity analytics. With a given clinical data set and a patient similarity context, we adopt a machine learning approach. First, we determine the set of features used to compute the similarity between patients, and determine the exact similarity metric based on the selected features where the outcome of each patient is considered. Then, we segment the patients into groups based on their similarities and characterize the groups with their unique characteristics. Downstream analytics can be performed on each group to discover insights for personalized care, e.g. local risk analysis and treatment efficacy analysis, which are beyond the focus of this paper. We describe the pipeline analytics in detail as follows.

Patient Similarity Determination

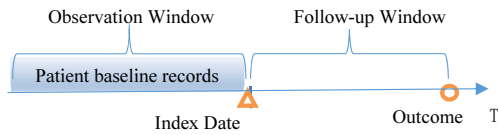


Figure 2 - Alignment of patient data with a specific similarity context

Patient Data Context Alignment

Given a clinical data set and a patient similarity context, we prepare the patient data as illustrated in Figure 2. For each patient, we identify an index date as the time when a clinical scenario is setting, e.g. when the diagnosis of type 2 diabetes is initially made, and use the records before the index date (observation window) to conduct the similarity analysis (those records after the index date and before the outcome date are ignored). The patient is then represented with a vector of features summarized from these clinical records. These features could incorporate the patient's information involving demographics, diagnosis, lab test, medication, and so on. In addition, we label the patient as positive or negative in terms of the outcome of interest (we now focus on the binary outcome only).

Similarity Feature Determination

Not all features are relevant to determine contextual patient similarities, and there are two typical ways to filter the relevant features from the candidates obtained from the previous step: 1) based on established domain knowledge (e.g. clinical guidelines), we identify the set of relevant features known as risk factors regarding the specific outcome; 2) we apply supervised feature selection methods [6] to automatically select the features relevant to the outcome. The analysis scenario determines which method to apply.

Similarity Metric Determination

Since the distance metric can be regarded as a measure of dissimilarity, similarity learning is closely related to metric learning. A few alternatives exist to measure the clinically similarity between two patients represented by vectors of the selected similarity features:

- Using the Euclidean distance between them or between their corresponding vectors with reduced dimensions by PCA (Principal Component Analysis) if there are too many similarity features.
- Using the distance between their predicted risk scores regarding the outcome of interest where we first develop a risk prediction model based on the selected features, and then applying the model to compute the risk score for each patient regarding the outcome. Patients with similar risk scores are considered similar.
- Using a learned Mahalanobis distance between them, which can automatically adjust the importance of each feature against the given outcome of interest. Formally, we represent a patient as a N-dimensional feature vector x where N is the number of identified relevant features. Let S be the set of equivalence constraints denoted by $S = \{(x_i, x_j) | x_i, x_j \text{ belong to the same outcome class}\}$ and D be the set of inequivalence constraints denoted by $D = \{(x_i, x_j) | x_i, x_j \text{ belong to the different outcome classes}\}$. Our goal is to learn a generalized Mahalanobis distance between patient x_i and patient x_j defined as:

$$d_A(x_i, x_j) = \sqrt{(x_i - x_j)^T A (x_i - x_j)}$$

Where A is positive semi-definite matrix and is designed by solving the optimization problem:

$$\min_{A \in R^{N \times N}} \sum_{x_i, x_j \in S} d_A^2(x_i, x_j) \quad \text{s.t.} \quad \sum_{x_i, x_j \in D} d_A^2(x_i, x_j) \geq 1$$

As a result of metric learning, we expect to keep pairwise vectors in S close and those in D separated away.

While a few algorithms [7] have been proposed to learn a Mahalanobis distance metric, in this study we have implemented three popular ones including linear discriminant analysis (LDA) [8], which projects the original feature vectors into a subspace that preserves the variance between class labels; large margin nearest neighbor (LMNN) [9], which learns a linear transformation of the input space where k nearest neighbor should have matching labels; and information-

theoretic metric learning (ITML) [10], which maximizes the differential entropy of a multivariate Gaussian subject to constraints on the associated Mahalanobis distance.

To determine which metric to use for computing patient similarities, we evaluate them using a nearest-neighbor based method. That is, for a specific metric, we compute the classification performance against the given target outcome using a KNN (K Nearest Neighbor) classifier that is built upon the metric. We consider a metric a better fit if it achieves the best classification performance.

Precision Cohort Finding

Patient Clustering

We apply agglomerative hierarchical clustering to group patients so that the patients within a group are contextually similar. The method starts with singleton clusters and proceeds by successively merging the two “closest” clusters at each stage. We customize the method by using the previously determined distance metric to determine the distances between clusters rather than using the conventional unsupervised distance metrics. With the learned distance metric, we expect that there could be a substantial divergence in the result of the targeted outcome (proportion of patients with a positive outcome) between the resultant groups.

One advantage of hierarchical clustering is the flexibility of determining the number of produced clusters based on the results of one running. For our purpose, we determine the number of reported clusters and evaluate the clustering performance by: 1) an internal clustering performance metric: the silhouette coefficient SH [11] which is defined as $SH = (b - a) / \max(a, b)$ where a is the mean distance between a patient and all other patients in the same group, and b is the mean distance between a patient and all other patients in the next nearest group. A SH near 1 indicates that the sample is far away from the neighboring clusters, and a SH greater than 0.2 is generally considered to be a fair clustering result; 2) An external metric to measure the group outcome disparity (OD) between the resultant groups which is defined as the difference between the maximal and minimal positive outcome rates of the groups. This metric reflects the effect of stratified risks among the groups, and the grouping result with higher OD is better for our purposes.

Patient Group Characterization

After obtaining the patient groups with stratified outcome results, our interest is to identify the characteristics of each group and understand the differences between groups. We address this by two means: 1) we compare the key feature differences between groups using statistical tests (a Kruskal-Wallis test for continuous features and Pearson chi-square test for categorical features.); 2) in order to identify the unique characteristics of groups, we build a decision tree that can differentiate the resultant M clusters where M classes of patients are labeled in accordance with their respective cluster memberships. For easier applicability, we further derive the explicit rules from the tree to interpret the group membership of a patient. In this work, we use C5.0 [12] to build a decision tree and convert it to rules.

Results-Risk Stratification of AF Patients

The CHA2DS2-VASc (CV) [13] score ranging from 0 to 9 has been widely recommended and used to identify AF patients with a high risk of IS ($CV \geq 2$) who need to be treated with oral anticoagulant (e.g. Warfarin) or radiofrequency ablation (RFA). However, it is still arguable that the CV score may not precisely capture the risk of particular AF patients from local populations. For example, there is a subgroup of patients with high CV but a low IS-occurrence rate for whom anticoagulation

may not be indicated. This is crucial because anticoagulants may have severe side effects, such as warfarin-related bleeding, and RFA incurs additional economic burden on patients. It therefore might be unnecessary to treat those patients with truly low risk. Thus, this study aims to apply our proposed approach to identify such AF patient subgroups, who have truly low risk but are misclassified as high-risk by high CV score.

Data Set

We use a data set from a cohort study for around 18,000 AF patients across China [14]. The collected data includes patients' structured baseline records (i.e. demographics, history, medication history) and clinical records (i.e. interventions, outcomes) during follow-up visits in a 3-year period. We are interested in studying the risk of IS during one year of follow-up by using the baseline features of patients, and also a selected population of 2,907 patients from the population (with an IS-occurrence rate is 4.6%) with the criteria as follows: 1) complete 12-month follow-up records are available for the patient, or follow-up records until the occurrence of an IS event within the 12-month follow-up period; 2) the patient either has no intervention (warfarin or RFA) until the end of 12 months of follow-up, or IS occurred before the intervention was started; 3) the patient's CV was ≥ 2 (i.e. the patient is considered high-risk). As the raw data has non-standard, missing and dirty values, we apply the same approach as our previous study [14] to automatically clean the data and impute the missing values. In the end, we have a data set with 132 input features and a binary feature for the outcome of having IS occurrence in 12 months of follow-up.

Furthermore, to validate the study results of analytics, we split the data into a derivation patient set (1,743 patients) and a validation patient set (1,164 patients) (60% and 40% of the population, respectively). The splitting strategy is to keep both the IS-occurrence rates and the CV score distributions the same between the two sets. We observe that the lowest IS-occurrence rate for the patients with $CV=2$ is approximately 2.5%, thus our objective is to discover a subgroup of patient with truly low risk of IS where the IS-occurrence rate should be lower than 2.5%.

Selecting Similarity Features

From the original data set with 132 input features, we first remove those that are relevant to IS occurrence but have strong correlation with known risk factors as defined by the CV score. We then automatically select the other relevant features using a filter-based method (SPSS modeler version 17). In other words, we select the top significant continuous features based on the p-value of using the F statistic and categorical features based on the p-value of using Pearson's χ^2 statistic ($p \leq 0.05$). Besides the four known CV features including prior CHF, prior IS, prior vascular diseases, and age, Table 1 lists the other 10 features selected as potential risk factors.

Table 1 – Similarity features selected from the data

Whether having a history of established coronary artery disease (ECAD)
Whether using drugs for ventricular rate control at the baseline (VRCD)
Whether there is a IS within the recent 5 years (IS5)
Whether there is a CHF within the recent 5 years (CHF5)
Whether there is a DM within the recent 10 years (DM10)
Whether statin were used to treat hyperlipidemia (Statin)
Total bilirubin at the baseline (TBIL)
Whether ACEI was used at the baseline (ACEI)

Left ventricular septum thickness on echocardiography at the baseline (IVS)
Left ventricular posterior wall thickness on echocardiography at the baseline (LVPW)

Determining the Similarity Metric

With the derivation data set where each patient is represented by 14 features above, we develop a few distance metrics to measure patient similarity (using scikit-learn 0.17): 1) EUCL. The Euclidean distance between the patient vectors; 2) LR_Score. The distance between the predicted risk scores using a logistic regression (LR) model trained on the derivation set (AUC of the derived LR is 0.72). 3) LDA. The learned Mahalanobis distance using LDA. 4) LMNN. The learned Mahalanobis distance using LMNN 5) ITML. The learned Mahalanobis distance using ITML. Table 2 reported their KNN classification performance on the validation set with the averaged results when KNN classifiers were trained on the derivation set with K set to 1, 3, 5, 7, and 9 respectively. While the overall F1 score is low for all metrics (this is due to our highly imbalance data set), LDA outperforms the other metrics.

Table 2 – KNN classification performance of different metrics.

	EUCL	LR_Score	LDA	LMNN	ITML
Average Precision	0.481	0.262	0.594	0.483	0.410
Average Recall	0.124	0.131	0.153	0.122	0.112
Average F1	0.132	0.151	0.194	0.132	0.134

Patient Grouping Results

Table 3 – Clustering results using different similarity metrics

	2 clusters			3 clusters		
	IS_occurrence rate	SH	OD	IS_occurrence rate	SH	OD
EUCL	3.8%, 6.6%	0.20	2.8%	3.8%, 5.3%, 9.4%	0.22	5.6%
LR_Score	3.4%, 13.8%	0.56	10.4%	1.6%, 4.6%, 13.8%	0.48	12.2%
LDA	2.9%, 10.2%	0.58	7.3%	1.5%, 4.1%, 10.2%	0.52	8.7%
LMNN	0%, 4.7%	0.65	4.7%	0%, 4.1%, 10.6%	0.24	10.6%
ITML	4.5%, 5.4%	0.32	0.9%	0%, 4.5%, 6%	0.30	6%

With all the considerations above, we decide to adopt the three clusters resultant from the clustering with LDA for downstream analysis. As summarized in Table 4, on the one hand, there is a stratified risk of IS among groups where group 1, 2 and 3 corresponds to low, medium, and high risk groups respectively. Particularly the lowest risk rate is 1.5% in group 1 which is even close to the patients with CV=1 (1.4% in our source data set). This implies a group of patients with truly low risk of IS. On the other hand, the IS-occurrence rates increase with the rising of the CV median values of groups. This to some extent verifies the rough validity of CHA2DS2-VASc in this local population.

Table 4 – Summary of the resultant patient groups

	Derivation Set			Validation Set		
	1	2	3	1	2	3
Proportion	34.7%	41.1%	24.2%	33.0%	49.4%	17.6%
IS- Rate	1.5%	4.1%	10.2%	1.6%	4.8%	9.3%
CV (Median)	3	4	6	3	4	6

Moreover, we validate the resultant clustering model against the validation set where each patient is assigned to one of three clusters based on his similarity with the center of each cluster. In this way, the patients are also divided into three groups as shown in Table 4. We observe that it approximately coincides with the result from the derivation set, and in particular, a precision cohort with truly low risk of IS at 1.6% is successfully identified too. Likewise, we also test the clustering model with LR_Score against the validation set because clustering with LR_score achieves a comparable result with using LDA on the derivation set. However, it fails to get a satisfactory result on the validation set where among the resultant three clusters, the lowest risk rate is 3.4% and there is an extreme small group with only 61 patients.

Group Characterization Results

Table 5 summarizes the key baseline characteristics that are significantly different among the groups from the derivation set where a p value ≤ 0.05 is considered statistically significant using Pearson chi-square test. Group 1 patients are the youngest and tend to have the lowest rates of comorbidities while group 3 patients are the oldest and have the highest rates of all comorbidities. The conditions of group 2 patients are in between group 1 and group 3 in terms of either comorbidities or medication taken or examination results. These group differences coincide with their varied IS-occurrence rates.

Table 5 – Baseline characteristics of the resultant groups

Group (patient count)	1 (605)	2 (716)	3 (422)	P value
Age (median)	71	74	77	<0.001
CHF	12%	34%	65%	<0.001
Prior IS	3%	16%	56%	<0.001
Vascular diseases	17%	22%	42%	<0.001
ECAD	10%	17%	32%	<0.001
CHF5	0%	9%	39%	<0.001
IS5	0%	4%	34%	<0.001
DM10	16%	8%	6%	<0.001
IVS (median)	9.7	9.8	10	<0.001
LVPW (median)	9.4	9.4	10	<0.001
VRCD	46%	80%	85%	<0.001
Statin	9%	25%	43%	<0.001

To further interpret the group characteristics, we build a decision tree using C5.0 (with a classification accuracy of 77% using SPSS modeler version 17) to classify these patients according to their cluster membership and derive a few rules to explicitly differentiate them. The resultant rules capture the unique characteristics of patients in different groups, and are easier to understand. In particular, we are the most interested in the patients in group 1 because they have the lowest IS-occurrence rate, and Table 6 lists the resultant 5 rules to characterize group 1 with high confidences (all are all above 85%). Due to the space limitation, we do not list the other grouping rules.

Table 6 - Rules to characterize the group 1 patients.

No	Rules	Confidence
1	Statin not used, no CHF in recent 5 years, no prior IS, and LVPW ≤ 7.9 mm	90.8%
2	Statin not used, no CHF in recent 5 years, no prior IS, age < 75 , and $8 \leq LVPW \leq 8.9$ mm	87.7%
3	VRCD not used, no CHF in recent 5 years, no prior IS, and age < 75	87.3%
4	Statin not used, no prior CHF, no prior IS, and age < 65	86.4%
5	VRCD not used, no prior CHF, no prior IS	85.5%

Table 7 - Validation result from the decision tree.

Group	1	2	3
Size (Percentage)	34.1%	41.2%	24.7%
IS-Occurrence Rate	2.4%	4.9%	6.6%
CV Score (Median)	3	4	6

We validate the developed decision tree by applying it to the validation set. As show in Table 7, the resultant three groups have the same CV median values with those from the clustering. To our interest, group 1 patients still have a low rate of IS-occurrence at 2.4% which is even lower than the patients with CV=2 (2.5% in our source data set). The results above support validity and stability of our approach.

Discussion

One issue for further investigation is if the group 1 patients that we identified with truly low risk are simply a subset of patient with the lowest CV score. To answer that, we compared the breakdown of CV distribution of group 1 patients between the clustering results on the derivation and validation sets, and the result shows that more than 50% of patients are actually have a CV>2 in both sets. This implies that CHA2DS2-VASc may not precisely fit this local population. In fact, prior hypertension and sex are not selected as our similarity features. This suggests a further systematic study on risk factor evaluation which should include all CHA2DS2-VASc factors and the novel feature candidates as in Table 1.

The critical part of our proposed methodology lies in determining an appropriate similarity metric for a specific context. While the outcome-driven learned metrics (including LR_Score and Mahalanobis distance metrics) generally outperform the unsupervised EUCL, it is still challenging to identify the most appropriate one from various outcome-driven metrics. The performance of these metrics may vary depending on the different data sets or different clinical scenarios. Furthermore, in order to cope with a large and sparse data set, we could adopt deep phenotyping [15] (which is not the focus of this study) to learn a set of latent features to compute patient similarity rather than use the selected raw features.

Conclusions

In this study, we developed an outcome-driven approach to identify groups of similar patients in terms of a particular clinical outcome. We validated the effectiveness of the approach by grouping AF patients with high risk of IS into three subgroups using a real-world data set and then identifying a precise group of patients with low risk of IS and their unique characteristics. This may help to better inform IS risk stratification in clinical guidelines. Further research would be necessary to verify the utility of novel risk factors identified. Subdividing high-risk patient groups may better target personalized care recommendations and improve patient outcomes. For example, a clinician could prescribe a treatment to a patient based on a treatment effectiveness analysis and comparison of the patient's characteristics against those of a precision cohort which includes similar patients. Our future work includes adopting two-stage clustering to handle a very large data set and incorporating temporal similarity features and other outcomes of interest.

References

- [1] S. Ather, L.E. Peterson, V.G. Divakaran, et al. Digoxin treatment in heart failure—unveiling risk by cluster analysis of DIG data. *Int J Cardiol.* 2011 Aug 4;150(3):264-9.
- [2] Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol.* 2014 Oct 28;64(17):1765-74.
- [3] Ebadollahi S, Sun J, Gotz D, Hu J, Sow D, Neti C. Predicting Patient's Trajectory of Physiological Data using Temporal Trends in Similar Patients: A System for Near-Term Prognostics *AMIA Annual Symposium Proceedings.* 2010;2010:192-196.
- [4] Sun J, Wang F, Hu J, Ebadollahi S. Supervised Patient Similarity Measure of Heterogeneous Patient Records. *ACM SIGKDD Explorations Newsletter*, 2012;14(1).
- [5] G.C. Fonarow, K.F. Adams, W.T. Abraham, C.W. Yancy, W.J. Boscardin. Risk Stratification for In-Hospital Mortality in Acutely Decompensated Heart Failure: classification and regression tree analysis. *JAMA*, Feb 2005, 293 (5).
- [6] Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 2003;3:1157-1182.
- [7] Kulis B. Metric Learning: A Survey. *Foundation Trends in Machine Learning* 2012;5 287-364.
- [8] Hastie T., Tibshirani R. Discriminant adaptive nearest neighbor classification. *IEEE Pattern Analysis and Machine Intelligence*, 1996;18(6):607-16
- [9] Weinberger KQ, Saul LK. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Journal of Machine Learning Research* 10 (2009) 207-244
- [10] Davis JV, Kulis B, Jain P, Sra S, Dhillon IS. Information-Theoretic Metric Learning. *Proceedings of the 24th International Conference on Machine Learning*, 2007.
- [11] Rousseeuw PJ. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Computational and Applied Mathematics*, 1987 20: 53–65.
- [12] Kuhn M, Johnson K. *Applied Predictive Modeling*, 2013
- [13] Lip GY, Nieuwlaar R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest.* 137 (2): 263–72.
- [14] Li X, Liu H, Du X, et al. Integrated Machine Learning Approaches for Predicting Ischemic Stroke and Thromboembolism in Atrial Fibrillation. *AMIA Annual Symposium Proceedings.* 2016.
- [15] Pivovarov R, Perotte AJ, Grave E, Angiolillo J, Wiggins CH5, Elhadad N. Learning probabilistic phenotypes from heterogeneous EHR data. *J Biomed Inform.* 2015 Dec;58:156-65.

Address for correspondence

Haifeng Liu. Email: liuhf@cn.ibm.com.

Biomedical Informatics and the Digital Component of the Exposome

Guillermo Lopez-Campos^{a,b}, Mark Merolli^c, Fernando Martin-Sanchez^{d,b}

^a Centre for Experimental Medicine, Queen's University of Belfast, Belfast, Northern Ireland, United Kingdom

^b Health and Biomedical Informatics Centre, The University of Melbourne, Parkville, Victoria, Australia,

^c School of Health Sciences, Swinburne University of Technology, Hawthorn, Victoria,

^d Environmental and Participatory Health Informatics (ENaPHI) Research Group, Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, NY, USA

Abstract

Biomedical informatics plays a key role in the development of precision medicine and other new technology-enabled health disciplines. In this context, the exposome (defined as the whole set of exposures on an individual) has become a relevant topic. Although most of the research work in the exposome area has been conducted around the physical and chemical world, we sustain that in an increasingly digitised world more attention should be paid to the digital component of the exposome derived from the interactions of individuals with the digital world. We define this "Digital Exposome" as 'the whole set of tools and platforms that an individual use and the activities and processes that an individual engages with as part of his/her digital life. In addition we support that biomedical informatics can and should lead research in this area.

Keywords:

Biomedical Technology; Environmental Exposure; Medical Informatics

Introduction

Advancements in recent years across multiple technological and scientific disciplines (including biomedical informatics), have produced an unprecedented increase in biomedical knowledge. The availability of new analytical tools (e.g. ultra-fast DNA sequencing) have facilitated the development of new approaches such as personalized medicine or precision medicine assisting the use and interpretation of individual genetic information for medical purposes. More recently, the term precision medicine has been popularised, thanks initially to a) the announcement of the Precision Medicine Initiative (PMI) in the US (known now as *All of Us* Research Program) and further b) with the allocation of funding for its development [1].

Precision medicine is often considered as a continuation of the previous personalised/stratified medicine efforts, and it should also be associated with other elements such as participatory medicine, lifestyle, behavior and the individual exposure to environmental factors [2]. This very ambitious initiative explicitly acknowledges that an individual's health status is the result of various complex interactions between their particular genetic make-up (genome) and their environment [3].

A key element in our current understanding of said individual environmental factors, is the concept of the exposome (coined but Wild in 2005). In his original definition, this author referred to the exposome as the whole set of exposures of an

individual since conception to death [4]. This definition has been revisited several times since [5,6]. The concept of the exposome has received substantial support both in Europe, the USA and Japan and has been fully embraced nowadays by an array of scientists and research funding agencies within the context of precision medicine. However, growing interest in environmental factors is neither new, nor uniquely associated with precision medicine. It has a long trajectory in epidemiology. In spite of this, the primary difference between those approaches and current perspectives on the exposome is that in most cases, epidemiology adopts a population perspective [7], whereas the exposome is aimed at studying environmental health at the individual level.

The alternate perspective we wish to consider in this light is that concurrently, our society is immersed within a digital revolution, living in a pervasive landscape of information and communication technologies (ICT), which are penetrating an increasing number of facets of everyday life, including our health. Internet connectivity is becoming so ubiquitous, with individuals increasingly spending more time online, driven by different motivations [8].

Therefore, with this in mind, we ponder the notion that society and individuals alike are living increasingly immersed in a digital world. Hence, it is both prudent and paramount that the concept of the exposome be reinforced to include its digital component. This perspective provides a fuller approach for biomedical research and potential future clinical purposes.

Health outcomes derived from digital environments and digital stimuli

The precision medicine movement builds off the fundamental paradigm shift, suggesting not only that personalised treatment is important, but also that the patient becomes the centre of decision-making, empowered and engaged in health management options [9]. With the rising infiltration of self-quantification devices, social media and other digital tools, it has become a relevant topic for health research along the last few years [10].

Concepts such as 'digital medicine' and 'digital health' have become greatly popular and describe a whole set of tools and approaches to health based on the digital revolution [11-13]. These terms cover a broad range of elements. In this context, the concept of exposure to different stimuli and Internet factors is not new and it has been used for different purposes since the early days of Internet [14,15].

In reference to impacts of immersion in digital environments, literature continues to appear in growing prevalence,

indicating that health outcomes may be precipitated in relation to use of various digital technologies. For example, social media use and their impacts on health outcomes of people managing chronic illness [16]. Whilst health effects are reported both positively and negatively in this context [17], the theory we reiterate in this manuscript is that regardless of outcome, exposure to various digital environments is a key component of the exposome of individuals.

Examples exist all the way through from basic website exposure, through to video games and social media as will be discussed.

Websites:

At a preliminary level, literature has gone so far as to show that browsing content of certain websites can be associated with health effects. This can be seen in the realm of literature commenting on correlation between web-based content and eating disorders (i.e. anorexia nervosa) [18,19].

Video Games:

Another area worthy of attention and of great interest for the study of digital health effects is engagement with video games. Gaming has garnered increasing attention in health and biomedical informatics through the inception of concepts such as 'gamification', with several instances of video game use and health outcomes reported [20-22]. This application highlights the power of the digital environment (and digital stimuli) to impact some element of health outcomes.

More specific examples of video game immersion have shown that the stimuli associated to recreational use of videogames can in-turn precipitate various effects, some deleterious, [23, 24] including the death of a patient due to cardiac failure associated with the stressed induced from gaming [25].

Social Media:

Social media have become an increasingly important part of everyday life, with literature commenting on social media's impact in health growing exponentially [26]. One area where it is clear that digital exposure can play a role in health related outcomes has been in the selection and use of different social media platforms as part of the management of chronic diseases [27]. A 2012 example indicated improved scores on stress, depression and anxiety scales after immersion in a virtual world created for the purposes of delivering mindfulness meditation training [28].

Further Exploring Digital Exposure – A Social Media Case Example

Here we focus on the topical area of social media, examining exposure of patients to different social media platforms based on previous works and experiences of the authors of this manuscript [13,14,24]. Using the case of social media, we present how different digital exposures may be associated with different health outcomes.

Social media or, peer-to-peer networks represent a layer of the digital environment in which participatory citizens are immersing themselves digitally as part of their health management. Empowered individuals are turning to such tools (i.e. social network sites, blogs, microblogs, wikis and virtual worlds) to source information about their conditions, connect with others, chat and share resources about their health, symptoms and treatments [10,16].

The influence of exposure to social media environments on one's health outcomes cannot be overlooked [29]. Previous research has successfully depicted the influence of online social networking on weight loss, smoking cessation, and depression, arguing that decisions to change health behaviours

are not only intrinsically motivated but also influenced by the effects of the individuals' online social network [29].

Based on our previous research, the same may be said for patient-reported health outcomes and the influence of social media based environments. Research by the authors of this manuscript [16,27] has qualified how social media use may impact health outcomes through the lens of what social media therapeutically afford the individual. This process has focussed on various health variables such as: cognitive health, social health, psychological health and physical health and the therapeutic affordances of self-presentation in the online world, connection to others, exploration of online information, narration of experiences with illness and through adapting online behaviour based on disease specific needs [16].

Underlying this process is the theoretical proposition that social media exposure impacts the individual. Our research suggests the following social media exposures and their postulated impact:

- *Time spent online (engagement)*: Literature review suggested a positive correlation between the number of times online social networking features are accessed, length of time spent on social networking and participation/engagement in health interventions [16]. Furthermore, statistical correlation also suggested that more frequent use of social media was positively linked to greater reported improvements in both psychological and social health variables in chronic pain management [27].
- *Interaction with others (support)*: People living with chronic diseases (i.e. arthritis, cancer, fibromyalgia, diabetes) have reported a strong sense of support received from social online environments. Exposure to online support through interaction with others has been noted to improve acceptance and validation of one's condition in complex regional pain syndrome and adaptive coping in HIV/AIDS for example [16] [30-32]. Other reports have also indicated exposure to peer support leads to improvements in empowerment, as well as increased participation in positive behaviours and activities [33].
- *Information access*: Information seeking is arguably still the most observed use of social media platforms, despite ability to connect and converse on these platforms [27]. Sourcing and improving disease-specific knowledge is reported as an important factor in self-management [16; 34; 35] and social media use has been shown to correlate to improved cognitive health surrounding one's condition [27].
- *Simulated interaction*: Virtual and augmented reality environments are becoming more apparent in social media and chronic disease literature. The virtual environment allows users to interact and navigate the virtual world anonymously using an avatar [28,35].

The Digital Exposome

In the recent years, interest in exposome-related research has increased enormously as it provides a framework to access and assess individual exposures relevant for health. In many cases, this access is being mediated by advances in the development, miniaturisation and integration of sensors, enabling the collection of individualised environmental data deemed to be important for individual health management. In this context a "Digital Phenotype" [36] has been previously described considering part of the digital footprint of an individual as a relevant element able to be related with

manifestations of disease. Simultaneously, individuals are becoming increasingly engaged with a digitalised environment which undoubtedly has an effect in every individual's health. Therefore there is a growing need to realise the timely relevance and importance of the digital subset of the exposome, the "Digital Exposome", as shown in Figure 1.

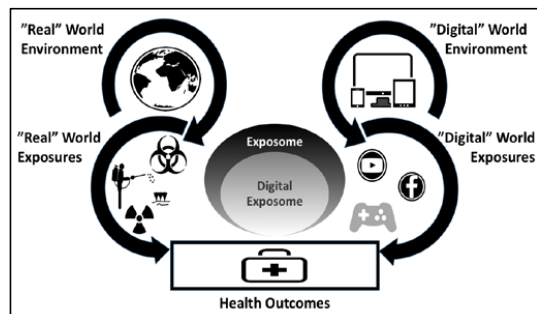


Figure 1– Both “real”(physical) and “digital” world exposures contribute to different individual health outcomes

In order to identify to what extent the concept of the digital component of the exposome has already been used in the literature, we carried out a Pubmed search using different relevant terms. Table 1.

Table 1– Results of a Pubmed search for different terms related with the digital exposome

Term	Total number of papers
Exposome	222
Exposome + Internet (MeSH)	0
Exposome + Software (MeSH)	1
Exposome + Digital Health	1
Exposome + Informatics	24

After conducting this search, it was clear that despite the relevance of the digital component of the exposome for health purposes, it has not as yet been sufficiently considered as a distinct element in the exposome literature.

Whereas in the physical environment the characterisation of an individual's exposome mostly relies in the use of surveys, sensors, geographical information systems, analytical tools and devices [37-40], the digital realm would require a completely different set of approaches reliant on the development of software able to identify those components of the digital exposome deemed relevant for health purposes. In clear contrast with the physical world exposome (where capturing exposure to the different elements quite often requires the development of individual devices), monitoring the digital exposome will require a novel approach. One in which software is developed to trace all digital activities that are relevant to the health of an individual. Some early examples of digital tracing already exist, such as in electronic commerce sites where personalised suggestions are made based on existing electronic

information of previous use of these online platforms (e.g.: Amazon, Netflix). Other attempts showed that although accessing to the individual digital exposome is to some extent technically feasible [41], it presents massive ethical challenges (e.g. data privacy) that require further elaboration [42].

The quantified-self (QS) movement aims to monitor different aspects of our daily lives and makes extensive use of digital tools [43,44]. Unlike the QS movement, where elements come together to engage with the digital environment to store, analyse and occasionally share this personal information for health and wellbeing purposes, the idea of self-monitoring our Internet use and other digital exposures in an integrated manner has not yet been sufficiently developed.

Biomedical Informatics and the Digital Exposome

The concept of the Exposome plays a critical role in the development of precision medicine. At the same time, the exposome, as a relatively new and strongly interdisciplinary area, poses new challenges for biomedical informatics [45]. The need for formal and systematic approaches to information processing in this context justifies a leading role for Biomedical Informatics within this new research avenue.

Most exposome-related research has been focused around the characterization of the “real-world” exposome, associated with physical, chemical and biological exposures[46,47]. Similarly, in early Precision Medicine studies and applications, “omics” approaches have played the central role, and so too has been the support provided by translational bioinformatics to them.

However, as reported in this manuscript, in modern societies where individuals are spending increasing amounts of time online or digitally connected, our digital exposures are becoming increasingly relevant to our health and therefore should be integrated with and considered as a relevant element of an individual's exposome within a Precision Medicine context. Notably, the role of biomedical informatics goes far beyond these initial approaches. As seen throughout this work, there exist many other areas where exposure information is derived from different online or digital experiences. In these cases biomedical informatics' role expands from the well known aspects of data management and analysis, to becoming an essential player in the generation of individual environmental data, with a particular emphasis on collecting and aggregating individual digital exposome data.

Therefore, further research work will be needed to characterize the digital component of the exposome, which we define as ‘the whole set of tools and platforms that an individual use and the activities and processes that an individual engage with as part of his digital life’. In many ways, the digital component of the exposome can be considered alongside the digital footprint of an individual. This concept of the “Digital Exposome” comes to complement the previously described “Digital Phenotype” in an holistic view of how the phenotype of an individual, and her health status, is defined by the complex interactions between the “real” (physical) world its digital component and the genetic component at an individualised level. Figure 2.

It is our aim to raise awareness about the need to monitorise individual exposome data and take into greater consideration the digital exposome as well as to highlight the key leading role that biomedical informatics should play if we are to foster research in this area to improve health outcomes. This could also have consequences in the design and content of biomedical informatics education and training programs.

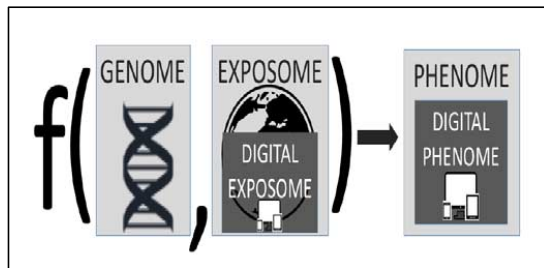


Figure 2—Phenotypes (including the digital phenotype) are the result of complex interactions between genome and exposome, including the digital exposome.

References

- [1] Precision Medicine, in, The White House - Office of the Press Secretary, Washington, 2015.
- [2] G.H. Lopez-Campos, V. Lopez-Alonso, and F. Martin-Sanchez, Is Precision Medicine different from Personalised Medicine? A Biomedical informatics perspective, *Stud Health Technol Inform* **202** (2014), 20-23.
- [3] National Research Council (U.S.). Committee on A Framework for Developing a New Taxonomy of Disease., *Toward precision medicine : building a knowledge network for biomedical research and a new taxonomy of disease*, National Academies Press, Washington, D.C., 2011.
- [4] C.P. Wild, Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology, *Cancer Epidemiol Biomarkers Prev* **14** (2005), 1847-1850.
- [5] C.P. Wild, The exposome: from concept to utility, *Int J Epidemiol* **41** (2012), 24-32.
- [6] G.W. Miller and D.P. Jones, The nature of nurture: refining the definition of the exposome, *Toxicol Sci* **137** (2014), 1-2.
- [7] P.J. Lioy and S.M. Rappaport, Exposure science and the exposome: an opportunity for coherence in the environmental health sciences, *Environ Health Perspect* **119** (2011), A466-467.
- [8] J.R. Anderson, L., Digital Life in 2025, in, Pew Research Center, 2014.
- [9] J.L. Fox, Obama catapults patient-empowered Precision Medicine, *Nat Biotechnol* **33** (2015), 325.
- [10] M. Swan, Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen, *J Pers Med* **2** (2012), 93-118.
- [11] E. Elenko, L. Underwood, and D. Zohar, Defining digital medicine, *Nat Biotechnol* **33** (2015), 456-461.
- [12] E. Topol, Digital medicine: empowering both patients and clinicians, *Lancet* **388** (2016), 740-741.
- [13] F.J. Martin-Sanchez and G.H. Lopez-Campos, The New Role of Biomedical Informatics in the Age of Digital Medicine, *Methods Inf Med* **55** (2016), 392-402.
- [14] M. White and S.M. Dorman, Receiving social support online: implications for health education, *Health Educ Res* **16** (2001), 693-707.
- [15] P.E. Kummervold, D. Gammon, S. Bergvik, J.A. Johnsen, T. Hasvold, and J.H. Rosenvinge, Social support in a wired world: use of online mental health forums in Norway, *Nord J Psychiatry* **56** (2002), 59-65.
- [16] M. Merolli, K. Gray, and F. Martin-Sanchez, Health outcomes and related effects of using social media in chronic disease management: a literature review and analysis of affordances, *J Biomed Inform* **46** (2013), 957-969.
- [17] M. Merolli, K. Gray, and F. Martin-Sanchez, Patient Participation in Chronic Pain Management Through Social Media: A Clinical Study, *Stud Health Technol Inform* **225** (2016), 577-581.
- [18] A.M. Bardone-Cone and K.M. Cass, What does viewing a pro-anorexia website do? An experimental examination of website exposure and moderating effects, *Int J Eat Disord* **40** (2007), 537-548.
- [19] M.J. Delforterie, J.K. Larsen, A.M. Bardone-Cone, and R.H. Scholte, Effects of viewing a pro-ana website: an experimental study on body satisfaction, affect, and appearance self-efficacy, *Eat Disord* **22** (2014), 321-336.
- [20] B. Cugelman, Gamification: what it is and why it matters to digital health behavior change developers, *JMIR Serious Games* **1** (2013), e3.
- [21] E. Gabarron, T. Schopf, J.A. Serrano, L. Fernandez-Luque, and E. Dorronzoro, Gamification strategy on prevention of STDs for youth, *Stud Health Technol Inform* **192** (2013), 1066.
- [22] M. Yates, A. Kelemen, and C. Sik Lanyi, Virtual reality gaming in the rehabilitation of the upper extremities post-stroke, *Brain Inj* **30** (2016), 855-863.
- [23] N. Ravaja, M. Turpeinen, T. Saari, S. Puttonen, and L. Keltikangas-Jarvinen, The psychophysiology of James Bond: phasic emotional responses to violent video game events, *Emotion* **8** (2008), 114-120.
- [24] C.V. Russoniello, K. O'Brien, and J.M. Parks, EEG, HRV and Psychological Correlates while Playing Bejeweled II: A Randomized Controlled Study, *Stud Health Technol Inform* **144** (2009), 189-192.
- [25] W. Grandner, J. Schwaiger, J. Seeber, and M.W. Dunser, Cyber war—when virtual fear of death turns into a real threat of life, *Int J Cardiol* **149** (2011), e16-17.
- [26] F.J. Grajales, 3rd, S. Sheps, K. Ho, H. Novak-Lauscher, and G. Eysenbach, Social media: a review and tutorial of applications in medicine and health care, *J Med Internet Res* **16** (2014), e13.
- [27] M. Merolli, K. Gray, and F. Martin-Sanchez, Therapeutic affordances of social media: emergent themes from a global online survey of people with chronic pain, *J Med Internet Res* **16** (2014), e284.
- [28] D.B. Hoch, A.J. Watson, D.A. Linton, H.E. Bello, M. Senelly, M.T. Milik, M.A. Baim, K. Jethwani, G.L. Fricchione, H. Benson, and J.C. Kvedar, The feasibility and impact of delivering a mind-body intervention in a virtual world, *PLoS One* **7** (2012), e33843.
- [29] R.V. Milani and C.J. Lavie, Health care 2020: reengineering health care delivery to combat chronic disease, *Am J Med* **128** (2015), 337-343.
- [30] C.L. Shigaki, K.L. Smarr, Y. Gong, K. Donovan-Hanson, C. Siva, R.A. Johnson, B. Ge, and D.R. Musser, Social interactions in an online self-management program for rheumatoid arthritis, *Chronic Illn* **4** (2008), 239-246.
- [31] M. Stelfox, B. Chaney, A.E. Barry, E. Chavarria, B. Tennant, K. Walsh-Childers, P.S. Sriram, and J. Zagora, Web 2.0 chronic disease self-management for older adults: a systematic review, *J Med Internet Res* **15** (2013), e35.
- [32] P.K. Mo and N.S. Coulson, Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups, *Int J Nurs Stud* **51** (2014), 983-993.
- [33] C.F. van Uden-Kraan, C.H. Drossaert, E. Taal, E.R. Seydel, and M.A. van de Laar, Participation in online patient support groups endorses patients' empowerment, *Patient Educ Couns* **74** (2009), 61-69.
- [34] L.S. Ruehlman, P. Karoly, and C. Enders, A randomized controlled evaluation of an online chronic pain self management program, *Pain* **153** (2012), 319-330.
- [35] C. Johnson, M. Feinglos, K. Pereira, N. Hassell, J. Blascovich, J. Nicollat, H.F. Beresford, J. Levy, and A. Vorderstrasse, Feasibility and preliminary effects of a virtual environment for adults with type 2 diabetes: pilot study, *JMIR Res Protoc* **3** (2014), e23.
- [36] S.H. Jain, B.W. Powers, J.B. Hawkins, and J.S. Brownstein, The digital phenotype, *Nat Biotechnol* **33** (2015), 462-463.
- [37] L. Faisandier, V. Bonnetterre, R. De Gaudemaris, and D.J. Bicout, Occupational exposome: a network-based approach for characterizing Occupational Health Problems, *J Biomed Inform* **44** (2011), 545-552.
- [38] G.M. Jaquez, C.E. Sabel, and C. Shi, Genetic GIScience: Toward a Place-Based Synthesis of the Genome, Exposome, and Behavior, *Ann Assoc Am Geogr* **105** (2015), 454-472.
- [39] M.J. Nieuwenhuijsen, D. Donaire-Gonzalez, M. Foraster, D. Martinez, and A. Cisneros, Using personal sensors to assess the exposome and acute health effects, *Int J Environ Res Public Health* **11** (2014), 7805-7819.
- [40] O.J. Robinson, X. Basagana, L. Agier, M. De Castro, C. Hernandez-Ferrer, J.R. Gonzalez, J.O. Grimalt, M.J. Nieuwenhuijsen, J. Sunyer, R. Slama, and M. Vrijheid, The Pregnancy Exposome: multiple environmental exposures in the INMA-Sabadell Birth Cohort, *Environ Sci Technol* (2015).
- [41] M. Harjuma, S. Saraniemi, S. Pekkarinen, M. Lappi, H. Simila, and M. Isomursu, Feasibility of digital footprint data for health analytics and services: an explorative pilot study, *BMC Med Inform Decis Mak* **16** (2016), 139.
- [42] V. Hordern, Data Protection Compliance in the Age of Digital Health, *Eur J Health Law* **23** (2016), 248-264.
- [43] M. Almalki, K. Gray, and F.M. Sanchez, The use of self-quantification systems for personal health information: big data management activities and prospects, *Health Inf Syst J* **3** (2015), S1.
- [44] M. Almalki, F. Martin Sanchez, and K. Gray, Self-Quantification: The Informatics of Personal Data Management for Health and Fitness in, Institute for a Broadband - Enabled Society (IBES), 2013.
- [45] F. Martin Sanchez, K. Gray, R. Bellazzi, and G. Lopez-Campos, Exposome informatics: considerations for the design of future biomedical research information systems, *J Am Med Inform Assoc* **21** (2014), 386-390.
- [46] S.M. Rappaport, D.K. Barupal, D. Wishart, P. Vineis, and A. Scalbert, The blood exposome and its role in discovering causes of disease, *Environ Health Perspect* **122** (2014), 769-774.
- [47] O. Robinson, X. Basagana, L. Agier, M. de Castro, C. Hernandez-Ferrer, J.R. Gonzalez, J.O. Grimalt, M. Nieuwenhuijsen, J. Sunyer, R. Slama, and M. Vrijheid, The Pregnancy Exposome: Multiple Environmental Exposures in the INMA-Sabadell Birth Cohort, *Environ Sci Technol* **49** (2015), 10632-10641.

Address for correspondence

Guillermo Lopez-Campos
drglopezcampos@gmail.com

Enhanced LexSynonym Acquisition for Effective UMLS Concept Mapping

Chris J. Lu^{a, b}, Destinee Tormey^{a, b}, Lynn McCreedy^{a, b} and Allen C. Browne^a

^a National Library of Medicine, Bethesda, MD, USA

^b Medical Science & Computing, Inc., Rockville, MD, USA

Abstract

Concept mapping is important in natural language processing (NLP) for bioinformatics. The UMLS Metathesaurus provides a rich synonym thesaurus and is a popular resource for concept mapping. Query expansion using synonyms for subterm substitutions is an effective technique to increase recall for UMLS concept mapping. Synonyms used to substitute subterms are called element synonyms. The completeness and quality of both element synonyms and the UMLS synonym thesaurus is the key to success in such applications. The Lexical Systems Group (LSG) has developed a new system for element synonym acquisition based on new enhanced requirements and design for better performance. The results show: 1) A 36.71 times growth of synonyms in the Lexicon (lexSynonym) in the 2017 release; 2) Improvements of concept mapping for recall and F1 with similar precision using the lexSynonym.2017 as element synonyms due to the broader coverage and better quality.

Keywords:

Natural Language Processing, Semantics, Unified Medical Language System

Introduction

Subterm substitution is a popular technique in query expansion. It is used to increase recall when no direct UMLS concept mapping is found through normalization. For example, no concept is found by direct mapping through normalization if the source vocabulary is “nasal deformity”. By substituting the subterm “nasal” for its synonym, “nose”, the UMLS concept [C0240547, Nose Deformity] is found, where “C0240547” is the concept unique identifier (CUI) and “Nose Deformity” is the preferred term in the UMLS. In this example, “nasal” and “nose”, which are used for substitution, are called element synonyms, while “nasal deformity” and “nose deformity” are the input term and expanded term, as shown in Figure 1. The normalized form of expanded terms is then used for concept mapping from UMLS synonyms.

Element synonyms are semantically equivalent terms (e.g. “nasal” and “nose” in the example above) used to identify subterms in the source vocabulary for substitution in UMLS concept mapping. This method increases recall by finding concepts for terms whose concept cannot be found by normalization or not even in the UMLS. For example, if “elderly” and “geriatric” are element synonyms, “elderly patients”, a term in the corpus (PubMed) but not in the UMLS, is mapped to the UMLS concept [C0199167, geriatric patients] by substituting “elderly” with its synonym, “geriatric”. The performance of this method relies on the quality and completeness of the element synonyms for a given UMLS thesaurus. The broader the coverage of the element synonyms,

the higher the recall. Commutativity and transitivity are two needed properties for quality element synonyms to preserve precision.

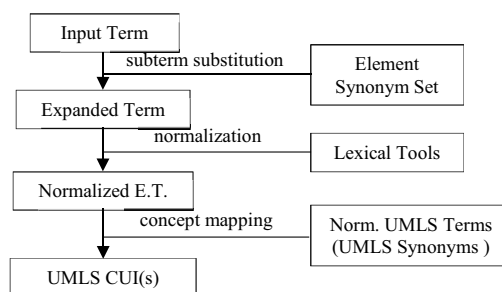


Figure 1 – Element Synonyms and Subterm Substitution in UMLS Concept Mapping

In this paper, we present a systematic approach to acquire a set of high quality element synonyms from the SPECIALIST Lexicon and UMLS Metathesaurus. The results show an improvement on recall and F1 with similar precision using this new acquired element synonym set for concept mapping.

Background

The 2016AA UMLS Metathesaurus of the National Library of Medicine (NLM), containing more than 3.25 million concepts and nearly 13 million unique concept names from over 190 source vocabularies, is one of the richest thesauri in the biomedical domain. UMLS concept mapping is used for managing knowledge in NLP applications including information retrieval (IR), document retrieval (DR), text classification, data mining, and decision support systems. Normalization is used as the initial step for UMLS concept mapping. All UMLS terms are processed through the Norm program in the Lexical Tools to normalize lexical variants, syntactic representation, and character encoding between ASCII and Unicode [1-2]. For example, “Behcet diseases”, “Behcet's disease, nos”, and “disease, Behcet” are UMLS synonyms because they represent the same concept. They have the same normalized term “behcet disease”. All UMLS terms are normalized and stored in the UMLS (MRXNS_ENG.RRF) with their associated concept(s). Terms having the same normalized form from input vocabulary (even if they are not in the UMLS Metathesaurus) can be mapped to UMLS concepts. For example, “disease, Behcet”, which is not a UMLS term, like other terms above, is mapped to [C0004943, Behcet Syndrome] through this normalization process.

Subterm substitution is used to find concepts for terms whose concept cannot be found through normalization. To increase

recall, strategies may use lexical or semantic information, or a combination of both. First, subterms can be substituted by lexically related variants, such as derivations. Derivations allow users to find closely related terms that may differ by part of speech (POS) for better recall [3-4]. For example, no CUI is found by direct mapping through normalization if the source vocabulary is “*perforated ear drum*”. By substituting the subterm “*perforated*” for its derivational variant, “*perforation*,” the UMLS concept is found [C0206504, Tympanic Membrane Perforation]. Second, subterm substitution by semantically equivalent terms (synonyms) improves recall [5-6]. Synonyms used for subterm substitution are terms that have the same meaning (concept) and are called element synonyms (such as “*nasal*” and “*nose*” from the example above). In practice, synonyms of synonyms are retrieved recursively (recursive synonyms) in such applications to increase recall. Third, subterms can be substituted by a combination of both lexical variants and synonyms [7-9]. These applications usually pre-generate all expanded terms and use them in a pool for concept mapping. The broader the coverage of the expanded terms, the better the recall for such approaches. Several works have used this strategy to find terms that the UMLS missed and improve recall [10-11]. This method of subterm substitutions generates many mapped concepts, including irrelevant concepts, and results in higher recall and lower precision. Ranking and filters, such as keyword match, frequency (TF-IDF), semantic types, concept distance and the longest lead-terms or end-terms, are used to improve the precision [12-13]. Other research has focused on different query expansion strategies by using UMLS Tools [14-15], MeSH [16-17] or their application systems [18-19] for effective UMLS concept mapping and information retrieval. Some research has explored the role of semantic similarity and semantic relatedness to similar and related terms having different UMLS concepts [20-21]. Prior to our work, there has been very limited effort devoted to acquiring element synonyms. Synonyms in the UMLS and Lexicon are two of most commonly used sources for element synonyms. However, several issues are found as described below.

UMLS synonyms with some restrictions, such as source vocabulary (MeSH), term length, and size of grams (usually unigram), were used as element synonyms for UMLS concept mapping in previous research [7-11, 14-16]. Three issues have been found in such approaches. First, UMLS synonyms are over-generated for element synonyms. For example, “*allergy drug*” and “*allergy medicine*” are UMLS synonyms, [C0013182, Drug Allergy] and considered as expanded terms. The concepts of these expanded terms can be found if their subterm, “*drug*” and “*medicine*,” are in an element synonym set. Slow runtime performance and computer resources are other concerns in practice when using the expanded terms of UMLS synonyms as element synonyms in subterm substitution, due to the large-scale size. Second, element synonyms must have properties of commutativity and transitivity for effective concept mapping. For example, “*ago*” is the abbreviation (ISO country code) for the country “*Angola*” and thus they are UMLS synonyms (with the same CUI, C0003023). However, “*ago*” is more often associated with another meaning, ‘earlier,’ and is not a synonym for “*Angola*” (lack of commutativity). In short, UMLS synonyms that represent broader or narrower concepts (such as “*adnexa*” and “*uterine adnexa*”), acronyms, abbreviations, POS ambiguity (e.g. “*mushroom*” is a synonym of “*Agaricales*” when its POS is a noun, but the meanings shift when its POS changes to a verb), terms with multiple CUIs, or the combination of the above, should be excluded from element synonyms. Acronyms with multiple CUIs cause a steep precision drop due to the large number of irrelevant mapped concepts in recursive subterm substitutions. For

example, the acronym of “*ER*” has more than 27 different expansions (concepts), such as “*emergency room*”, “*efficacy ratio*”, “*eye research*”, etc. Third, element synonyms may be single words (unigrams) or multiwords (words with spaces). Terms (multiwords) are used in more sophisticated systems as element synonyms to gain better recall [5-11]. For example, no concept is found for “*zona vaccine*”. By substituting “*zona*” for its multiword synonym, “*herpes zoster*”, the UMLS concept (C1720918) is found. On the other hand, longer terms introduce more noise rather than improving the performance. For example, “*herpes zoster infection*” a UMLS synonym of “*zona*” should not be used as an element synonym. To our best knowledge, there is no study on how many grams should be used for element synonyms.

In addition to UMLS synonyms, LexSynonyms are also commonly used as element synonyms for UMLS concept mapping in NLP. They are recorded in the format of synonym pairs (sPairs) with POS information and distributed with the SPECIALIST Lexicon. Two synonym records (sRecords) are generated by an sPair because sPairs are bi-directional. They are in the format of [synonym-1|POS-1|synonym-2|POS-2]. In most applications, they are integrated with lexical variants to generate expanded equivalent terms for concept mapping in MetaMap [7], MMTx [8], and Sophia [9]. LexSynonyms were originally collected as a set in the early 1990s and maintained manually by LSG linguists based on users’ requests. A rather static size of this synonym set is observed: only 142 sRecords were added between 2004 (5,056) and 2016 (5,198). Thus, we developed a systematic approach to acquire lexSynonyms as a standalone set of element synonyms with greater coverage and better quality for more effective UMLS concept mapping and NLP applications that use synonym retrieval.

Approaches

Synonyms can be categorized into two types: cognitive synonyms and near-synonyms. Cognitive synonyms have fewer meaning differences with greater interchangeability, while near-synonyms lack these. Cognitive synonyms match the characteristics of element synonyms well for effective performance (recall and precision) because they have two properties, commutativity and transitivity. Commutativity, $(x = y) \rightarrow (y = x)$, preserves the naturalness of bi-direction of sPairs. For example, if “*joy*” is a cognitive synonym of “*happy*”, then “*happy*” is a cognitive synonym of “*joy*”. Transitivity, $((x = y) \text{ and } (y = z)) \rightarrow (x = z)$, preserves the precision in recursive synonym applications. For example, if “*happy*” is a synonym of “*joy*”, and “*joy*” is a synonym of “*enjoy*”, then “*happy*” is a synonym of “*enjoy*”. These two properties are necessary conditions of quality element synonyms for subterm substitutions in concept mapping. However, they are missing in most synonym sets used in NLP. They are required for lexSynonym acquisition in our new system to ensure the effective UMLS concept mapping: lexSynonyms must be cognitive synonyms.

To acquire a thorough synonym set, UMLS synonyms are chosen as source candidates in this project. UMLS synonyms are UMLS strings (element terms and expanded terms) with the same concept (CUI). They are grouped and represented as a key-value collection in a synonym class (sClass). Namely, the key is the CUI while the value is the list of all terms with the same CUI in the UMLS Metathesaurus. This is the common way of retrieving UMLS synonyms. The derived UMLS sClass is further enhanced through the integration of the Lexicon. The Lexicon includes additional information needed for resolving the NLP issues mentioned above, such as POS, inflections, acronyms, abbreviations, etc. First, a lexical entry must be a word (single word or multiword) with a special unit of meaning

in itself [22-23]. The Lexicon is used as the source vocabulary to filter element synonyms: terms in the sClass that are not in the Lexicon, such as non-word phrases, are removed. For example, expanded terms of UMLS synonyms “allergy drug” and “allergy medicine” are removed to resolve the issue of over-generation, while “herpes zoster infection” is removed to resolve the issue of n-grams because none of them are in the Lexicon (do not meet the requirements of LexMultiwords) [23]. As discussed before, recall of concept mapping will not decrease because, “drug” and “medicine”; “zona” and “herpes zoster”, are terms in the Lexicon and used as element synonyms. Second, the POS information from the Lexicon is added to the sClass to resolve the POS ambiguity issues. Third, terms that are acronyms or abbreviations in the Lexicon are removed to preserve precision. Fourth, synonyms in the sClass need to be verified by experts to ensure they meet the requirements of commutativity and transitivity. Finally, the verified sClass is further processed into sPairs and sRecords to compose the element synonym set. All synonymous terms from the Lexicon (lexSynonyms) are acquired using this approach.

Implementation

A standalone lexSynonym set is established by collecting all synonymous terms in the Lexicon based on the above requirements and approaches. LexSynonyms are acquired from three types of sources: the Lexicon, the UMLS, and NLP projects. They are described as follows.

Lexicon-Sourced Synonyms – Nominalizations with EUI

Nominalizations are cognitive synonyms with the adjectives and/or verbs from which they are derived. They are recorded in the Lexicon and can be retrieved automatically to generate lexSynonyms. Additional information, the entry unique identifier (EUI) of the lexical record, is added to the associated sPair for downstream NLP processing. For example, the sPair of [ability|noun|able|adj|E0006490] is generated from the lexical record (E0006490). As shown in Figure 2, the noun of “ability” is the nominalization of the adjective, “able”.

```
{base=ability
entry=E0006490
  cat=noun
  variants=reg
  variants=uncount
  compl=pphr(of,np)
  compl=infcomp:arbc
  nominalization_of=able|adj|E0006510
}
```

Figure 2 – Lexical Record of C0011065, ability

UMLS-Sourced Cognitive Synonyms with CUI

The Lexicon and UMLS Metathesaurus are used to retrieve more synonymous lexicon terms as follows. First, all English terms from the UMLS (MRCONSO.RRF) with the same EUI are retrieved. Second, concepts of chemicals and drugs are removed due to limited resources and application domains. The semantic type indexes (STIs) of chemicals and drugs are used as filters through the mapping from CUI to STI (MRSTY.RRF). Third, terms having the POS of noun, verb and adjective with inflections of base in the Lexicon are retrieved. This step eliminates inflectional variants, illegal POSs, and non-word phrases from the UMLS synonyms. Fourth, terms that are acronyms or abbreviations in the Lexicon are removed. Fifth, terms with the same CUI are stored in an sClass with the CUI as the key and a list of terms as the value. The associated EUI is added to each term in the list for the computer to reference lexical records for needed information. Sixth, terms that are spelling variants (spVar) or nominalizations of other terms in

the same sClass are removed to save manual tagging time because they can be generated automatically later (in step nine). Seventh, sClasses with only one term are removed because they do not have synonyms. Eighth, UMLS preferred terms are added to sClasses for concept identification by LSG linguists when validating if terms (synonym candidates) are cognitive synonyms of the sClass. Ninth, spVars and nominalizations of validated synonym candidates are added back to the sClass. Tenth, tagged sClasses are used to generate the sPairs and sRecords with POS and source information (CUI, EUI and NLP). For example, “death”, “dead”, “deceased” and “die” are base forms with qualified POSs in the Lexicon, have the same CUI (C0011065), and are not chemicals, drugs, acronyms, or abbreviations. They are thus synonym candidates and are gathered in a candidate sClass as shown in Figure 3. Among the synonyms, “die” is related by nominalization to “death” (E0020918), and is thus removed. This is the candidate sClass sent to LSG linguists for validation. Cognitive synonyms are tagged as “Y” while near-synonyms are tagged as “N”. The nominalizations, “deadness” (E0020885) from “dead” (E0020877) and “die” from “death”, are added back into the sClass automatically. The final sClass is composed of 5 synonyms, generating 10 (bidirectional) sPairs, and results in 20 synonym records (sRecords) in the lexSynonym set, as shown in Figures 4 and 5 respectively.

```
#SYNONYM_CLASS|C0011065|Cessation of life
noun|E0020918|death|
adj|E0020877|dead|
adj|E0020990|deceased|
verb|E0022536|die|
```

Figure 3 – Example of Candidate sClass: C0011065

```
#SYNONYM_CLASS|C0011065|Cessation of life
noun|E0020918|death|Y
adj|E0020877|dead|Y
adj|E0020990|deceased|Y
verb|E0022536|die|nom
noun|E0020885|deadness|nom
```

Figure 4 – Example of Final sClass: C0011065

```
deadness|noun|dead|adj|C0011065
deadness|noun|death|noun|C0011065
deadness|noun|deceased|adj|C0011065
deadness|noun|die|verb|C0011065
dead|adj|deadness|noun|C0011065
dead|adj|death|noun|C0011065
dead|adj|deceased|adj|C0011065
dead|adj|die|verb|C0011065
death|noun|deadness|noun|C0011065
death|noun|dead|adj|C0011065
death|noun|deceased|adj|C0011065
death|noun|die|verb|C0011065
deceased|adj|deadness|noun|C0011065
deceased|adj|dead|adj|C0011065
deceased|adj|death|noun|C0011065
deceased|adj|die|verb|C0011065
die|verb|deadness|noun|C0011065
die|verb|dead|adj|C0011065
die|verb|death|noun|C0011065
die|verb|deceased|adj|C0011065
```

Figure 5 – Example of sRecords: C0011065

NLP Project-Sourced Cognitive Synonyms

Synonyms from NLP projects can be processed by similar steps to those described above, then added into lexSynonyms. For the 2017 release, we processed synonyms from Lexical Variants Generation (LVG). Duplicated synonyms of the previous two sources are removed from the candidate list without further process. Others are converted to sPair candidates computationally, reviewed by LSG linguists, and added to the lexSynonym set with POS if they are cognitive sPairs and in the Lexicon. “NLP_XXX” is used as the source information for the NLP project “XXX”. For example, “NLP_LVG” is marked as the source for synonyms from the LVG. The NLP project-

sourced synonyms provide two important features of extendibility and compatibility. First, users are able to extend the synonym set by adding domain/project specific synonyms. Second, it preserves the same result for the specific NLP project (LVG) users when forward compatibility is required.

Results, Tests, Discussions and Applications

As a result, 22,779 sClasses and 58,134 synonym candidates are retrieved from the UMLS source type (2016 AA UMLS Metathesaurus and 2016 Lexicon). Cognitive synonyms from this candidate list are used to generate 118,468 sRecords. In addition, 67,584 sRecords from Lexicon nominalizations and 4,792 sRecords from NLP_LVG are generated, respectively. All sRecords from these resources are combined into the lexSynonym set and distributed in the 2017 release of the Lexicon. The results show a growth of 36.71 times from 2016 to 2017 release through this new approach (Table 1).

Table 1 – Growth for LexSynonyms 2016 to 2017

Year	CUI	EUI	NLP	Total
2016	0	0	5,198	5,198
2017	118,468	67,584	4,792	190,844

A model is established to measure the performance of using the lexSynonym.2017 for UMLS concept mapping through the Sub-Term Mapping Tools (STMT). STMT applies a real-time subterm substitution algorithm for UMLS concept mapping with the configurable options of choosing element synonyms and UMLS release. The UMLS-CORE project assigned CUI(s) to terms (13,076) that are within the top 95% usage and mappable to SNOMED CT [5]. 2,755 of these terms (with 2,756 CUIs) without mapped concepts in UMLS.2016AB through normalization are used as the gold standard for this test. Five normalized element synonym sets are configured in STMT for comparison. The default STMT element synonym set is comprised of high quality synonyms for subterm substitution to improve recall (25%). They are validated cognitive synonyms from sources of British English, Greco-Latin, acronyms, abbreviations, Emergency Care Research Institute (ECRI), etc. [6]. Results are shown in Table 2: 1) recall is increased over 10% from lexSynonym 2016 to 2017 due to broader coverage (from 5K to 150K). Also, the precision is increased due to better quality. 2) recall and F1 are further improved about 5% and 0.05 while precision is about the same (-0.03%) by adding 2017 lexSynonyms to the STMT synonym set. The set of lexSynonym.2017 contains 5,872 (~75%) normalized synonyms in the STMT synonym set. Adding the previous lexSynonyms (2016) to STMT offers no improvement.

Table 2 – Test Result for Terms without Mapped Concepts

Synonym Set	N. Size**	Prec.	Recall	F1
STMT	7,873	66.16%	25.04%	0.3633
LS.2016*	5,070	42.86%	0.33%	0.0065
LS.2017	149,912	71.04%	10.41%	0.1816
STMT+LS.2016	12,681	65.87%	25.07%	0.3632
STMT+LS.2017	151,913	66.13%	30.04%	0.4132

* LS: LexSynonym Set, **N.: Size of Normalized Synonym Set

Due to limited resources, about 1/3 of synonym candidates (20,566 out of 58,134) have so far been tagged. The properties of commutativity and transitivity of lexSynonyms are ensured by nominalization (Lexicon-sourced) or by linguists' tags. 92.20% of synonym candidates are tagged as "Y". The size of the UMLS-sourced lexSynonym is about 0.64% of the size of the UMLS synonyms in English. Accordingly, the size of lexSynonyms will be about 2% of the UMLS synonyms when

the tagging process is completed. LexSynonyms thus yield a much smaller, more manageable set to be used as element synonyms. In addition, synonyms from other NLP projects, such as UMLS-CORE and STMT, can be further processed and added to the lexSynonyms. Recall is expected to be further improved as the size of element synonyms grows while the precision is preserved by the properties of cognitive synonyms.

We utilized lexSynonyms as element synonyms in NLP applications (Lexical Tools) to retrieve synonyms. Synonyms, POS, and source information are provided in the outputs of synonym features of Lexical Tools. A sophisticated algorithm is implemented as follows in the recursive synonym flow component to preserve precision. First, only synonyms with the same CUI are retrieved recursively if the source type is CUI. Second, all synonyms are retrieved recursively if the source type is EUI. Third, synonyms from the same NLP projects are retrieved recursively if the source type is NLP. In addition, the synonym source option (-ks) is implemented to allow users to restrict the results by source type (CUI, EUI, NLP), or any combination of the above. These new features provide needed information to preserve precision for downstream NLP processing. For example, the five synonyms of "die" are retrieved from the synonym feature (-f:y) in Lexical Tools. The source information is also included. As shown in Figure 6, "dead", "deadness", "death" and "deceased" are from the source of UMLS with CUI of [C0011065], while "expire" is from source of NLP (project LVG). The POS information is included in the outputs of the Lexical Tools. "Terminate", a synonym of "expire" from the resource of NLP_LVG, is retrieved when the recursive synonym feature (-f:r) is used in the Lexical Tools, as shown in Figure 7. The last two fields of the last line in Figure 7 show the source type (NLP_LVG) and the recursive history (yy, means synonym of synonym). Thus, project specific non-cognitive sPairs, "dead" and "terminate", can be distinguished by the different types of sources (CUI vs NLP) to preserve the precision in recursive synonyms.

```
die|verb|y|dead|adj|C0011065
die|verb|y|deadness|noun|C0011065
die|verb|y|death|noun|C0011065
die|verb|y|deceased|adj|C0011065
die|verb|y|expire|verb|NLP_LVG
```

Figure 6 – Synonyms of "die" from Lexical Tools

```
die|verb|r|dead|adj|C0011065|y
die|verb|r|deadness|noun|C0011065|y
die|verb|r|death|noun|C0011065|y
die|verb|r|deceased|adj|C0011065|y
die|verb|r|expire|verb|NLP_LVG|y
die|verb|r|terminate|verb|NLP_LVG|yy
```

Figure 7 – Recursive Synonyms of "die" from Lexical Tools

Conclusion

We have demonstrated the usefulness of the general concept of element synonyms as well as the Lexicon-specific type of element synonyms, lexSynonyms, in concept mapping. A systematic and maintainable approach is used to acquire higher quality lexSynonyms through the use of the Lexicon. Issues of over-generation and n-grams are resolved by restricting UMLS synonyms that are base forms with noun, verb, and adjective POS in the Lexicon, and removing chemicals and drugs. Terms that are acronyms or abbreviations are removed to avoid a drop in precision. Synonym candidates in the sClass that do not match the properties of commutativity and transitivity are tagged by the linguists as invalid to resolve near-synonym issues. POS is added to sPairs automatically through a Lexical records mapping by using EUIs in the sClass during the generation process. The information of source with unique identifier (CUI, EUI, and NLP) is also included. This

information is vital for downstream NLP applications to preserve precision especially when recursive synonyms are used. As a result, a thorough set of element synonyms is generated. LexSynonyms are expected to grow with the Lexicon and UMLS Metathesaurus for better coverage through this system. This approach is generic for element synonym acquisition and can be applied to other corpora, vocabularies, or synonym thesauri. The generated lexSynonyms are used in the Lexical Tools with enhanced recursive algorithms to provide better usage of the synonym related features for NLP applications. We believe the impact of better quality and broader coverage for lexSynonym acquisition in the Lexicon for effective UMLS concept mapping will improve the precision, recall, and naturalness of NLP applications. The set of lexSynonyms is distributed in the 2017 release of SPECIALIST Lexicon with UMLS by NLM via an Open Source License agreement.

Acknowledgements

This research was supported by the Intramural Research Program of the NIH, National Library of Medicine. The authors would like to thank Dr. Kin Wah Fung, Dr. Marcelo Fiszman, Guy Divita, Willie Rogers, James Mork and Francois-Michel Lang for their valuable discussions and suggestions.

References

- [1] A.T. McCray, S. Srinivasan, A.C. Browne. Lexical Methods for Managing Variation in Biomedical Terminologies. In proceedings of the 18th Annual Symposium on Computer Applications in Medical Care, 1994, 235-239.
- [2] C.J. Lu and A.C. Browne. Converting Unicode Lexicon and Lexical Tools for ASII NLP Applications. In proceedings of AMIA Annual Symposium, Oct. 22-26, 2011, 1870.
- [3] C.J. Lu, D. Tormey, L. McCreedy, A.C. Browne. A Systematic Approach for Automatically Generating Derivational Variants in Lexical Tools Based on the SPECIALIST Lexicon. IEEE IT Professional Magazine, May/June, 2012, 36-42.
- [4] C.J. Lu, D. Tormey, L. McCreedy, A.C. Browne. Generating SD-Rules in the SPECIALIST Lexical Tools - Optimization for Suffix Derivation Rule Set. HealthInf 2016, Feb. 21-23, 2016, Vol. (5), 353-358.
- [5] K.W. Fung and J. Xu. An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT. JAMIA 2015, 22: 649- 658.
- [6] C.J. Lu, and A.C. Browne. Development of Sub-Term Mapping Tools (STMT). In Proceedings of AMIA Annual Symposium, Nov. 3-7, 2012, 1845.
- [7] A.R. Aronson. The Effect of Texture Variation on Concept based Information Retrieval. In proceedings of AMIA Annual Symposium, 1996, 373-377.
- [8] G. Divita, T. Tse, L. Roth. Failure Analysis of MetaMap Transfer (MMTx). In proceedings of MedInfo 2004, Sept. 7-11, 2004, 763-767.
- [9] G. Divita, Q.T. Zeng, A.V. Gundlapalli, et al. Sophia: A Expedient UMLS Concept Extraction Annotator. In proceedings of AMIA Annual Symposium, Nov. 15-19, 2014, 467-476.
- [10] W. Hole, S. Srinivasan. Discovering Missed Synonymy in a Large Concept-Oriented Metathesaurus. In proceedings of AMIA Annual Symposium, Nov. 4-8, 2000, 354-358.
- [11] K.C. Huang, J. Geller, M. Halper, J.J. Cimino. Piecewise Synonyms for Enhanced UMLS Source Terminology Integration. In proceedings of AMIA Annual Symposium, Nov. 10-14, 2007, 339-343.
- [12] T.C. Eskridge, A. Granados, A.J. Cañas. Ranking Concept Map Retrieval in the CmapTools Network. In proceedings of the 2nd International Conference on Concept Mapping, 2006. Vol. 1, 477-484.
- [13] H.C. Wu and R.W.P. Luk. Interpreting TF-IDF Term Weights as Making Relevance Decisions. ACM Transactions on Information Systems, June, 2008, Vol. 26, No. 3, Article 13.
- [14] N. Griffon, W. Chebil, L. Rollin, G. Kerdelhue, B. Thirion, J.F. Gehanno, and S. J. Darmoni. Performance evaluation of unified medical language system®'s synonyms expansion to query PubMed. BMC Medical Informatics and Decision Making 2012, 12(1): 12.
- [15] K. Lu, X.M. Mu. Query Expansion Using UMLS Tools for Health Information Retrieval. J. of the American Society for Information Science and Technology, 2009, 46 (1), 1-16.
- [16] M.C. Díaz-Galiano, M.T. Martín-Valdivia, L.A. Ureña- López. Query expansion with a medical ontology to improve a multimodal information. Comp. in Bio. and Med. 2009, 39(4): 396-403.
- [17] M. Berardi, M. Lapi, P. Leo, and C. Loglisci. Mining Generalized Association Rules on Biomedical Literature. IEA/AIE, 2005, 500-509.
- [18] P. Srinivasan. Query Expansion and MEDLINE. J. of Information Processing & Management. 1996, Vol. 32, No. 4, 431-443.
- [19] Q.T. Zeng, D. Redd, T. Rindfleisch, J. Nebeker. Synonym, Topic Model and Predicate-Based Query Expansion for Retrieving Clinical Documents. In proceedings of AMIA Annual Symposium, Nov. 3-7, 2012, 1050-1059.
- [20] T. Pedersen, S.V.S. Pakhomov, S. Patwardhan, C.G. Chute. Measures of Semantic Similarity and Relatedness in the Biomedical Domain. J. of Biomedical Informatics, 2006, 40(3), 288-299.
- [21] S.V.S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, G. Melton. Semantic Similarity and Relatedness between Clinical Terms: An Experimental study. In proceedings of AMIA Symposium, Nov. 13-17, 2010, 572-576.
- [22] C.J. Lu, D. Tormey, L. McCreedy, A.C. Browne. Multiword Frequency Analysis Based on the MEDLINE N-Gram Set. In proceedings of AMIA Annual Symposium, USA, Nov. 12-16, 2016, 1488.
- [23] C.J. Lu, D. Tormey, L. McCreedy, A.C. Browne. Generating A Distilled N-Gram Set: Effective Lexical Multiword Building in the SPECIALIST Lexicon. HealthInf 2017, Porto, Portugal, Feb. 21-23, 2017, Vol. 5, 77-87.

Address for correspondence

Dr. Chris J. Lu, chlu@mail.nih.gov

NIH/NLM/LHC/CgSB/MSC

8600 Rockville Pike, Bldg. 38-A, B1N-28R

Bethesda, MD 20894

A Validated Risk Model for 30-Day Readmission for Heart Failure

Satish M. Mahajan^{a,b}, Prabir Burman^c, Ana Newton^d, Paul A. Heidenreich^{a,e}

^a Veterans Health Administration, Palo Alto, California, USA,

^b School of Nursing, University of California Davis, Davis, California, USA,

^c Department of Statistics, University of California Davis, Davis, California, USA,

^d College of Public Health, University of San Francisco, San Francisco, California, USA,

^e Department of Cardiology, Stanford University, Stanford, California, USA

Abstract

One of the goals of the Precision Medicine Initiative launched in the United States in 2016 is to use innovative tools and sources in data science. We realized this goal by implementing a use case that identified patients with heart failure at Veterans Health Administration using data from the Electronic Health Records from multiple health domains between 2005 and 2013. We applied a regularized logistic regression model and predicted 30-day readmission risk for 1210 unique patients. Our validation cohort resulted in a C-statistic of 0.84. Our top predictors of readmission were prior diagnosis of heart failure, vascular and renal diseases, and malnutrition as comorbidities, compliance with outpatient follow-up, and low socioeconomic status. This validated risk prediction scheme delivered better performance than the published models so far (C-Statistics: 0.69). It can be used to stratify patients for readmission and to aid clinicians in delivering precise health interventions.

Keywords:

Patient Readmission; Models, Theoretical; Heart Failure

Introduction

Based on the Precision Medicine Initiative (PMI) launched in the United States in 2016, the director of the National Institutes of Health (NIH) convened a working group that defined Precision Medicine as “an approach to disease treatment and prevention that seeks to maximize effectiveness by taking into account variability in genes, environment, and lifestyle” [1]. The group also identified development of quantitative estimates of risk for a range of diseases. The initiative further established that the need for pioneering efforts to merge, integrate, and analyze data from various resources including Electronic Health Records (EHR) [1]. Therefore, we sought to develop a use case for the goals established by the PMI by considering 30-day hospital readmissions for patients with Heart Failure (HF).

According to Fida and Pina, HF is the most expensive Diagnosis Related Groups (DRG) diagnosis for hospitalizations and the most frequent diagnosis for 30-day readmissions [2]. As a part of the American Heart Association's Policy Statement on forecasting the future of cardiovascular diseases in the United States, Heidenreich et al. provide projections for the 2010 to 2030 timeframe for HF that forecast increases of 215% and 80% in direct and indirect costs respectively, and a corresponding increase in HF prevalence by 25% [3]. Thus, reducing the readmission rate,

especially for HF, is important to quality-conscious and fiscally responsible hospitals in delivering precise health interventions.

A data-driven risk assessment approach would facilitate identification of patients at potentially high risk for HF readmissions [4]. Towards this goal, many researchers have proposed predictive models for worsening symptoms, mortality, and/or readmissions. However, there is no consistency across the models in terms of their use of readmission timeframe, data sources, algorithms, and demonstration of predictive power. Some models used either registry- or trial-based clinical data [4-8], while others used hospital-based or Medicare claims-based administrative data [9-12]. Recently, researchers suggested examining data surrounding patients' overall health [13-15] to elicit better predictivity from such models. We propose a systematic combination of structured data from clinical, administrative, and psychosocial domains to understand the predictivity of each group on the 30-day readmission. We hypothesize that a model with the incremental inclusion of each domain would have a better prediction for the 30-day readmission rate than domain-specific models under sensitivity analysis.

Methods

Patient Population

We used data from the EHR at the Veterans Affairs Palo Alto Health Care System, Palo Alto, California (Vista) to derive and validate the predictive models for this observational retrospective cohort study. The data between October 1, 2005 and September 30, 2013 was used to build the cohort identified by using International Classification of Diseases version 9 – Clinical Modification (ICD-9-CM) codes for heart failure as a principal discharge diagnosis: 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9 [9; 13].

Primary Outcome

We adhered to the literature definition of 30-day readmission that indicates an all- or any-cause readmission within 30 days of discharge from the index hospitalization for the comparison of our results with the other published result. We excluded hospitalizations with a length of stay less than one day and elective hospitalizations from the readmission episode. If the patient had multiple episodes of 30-day readmissions, only the last episode of such readmissions was considered. We thus

produced a statistically independent and mutually exclusive sample of patients across the two classes of the cohort.

Predictors

Predictor variables were collected from the EHR with the understanding that they were routinely recorded and available across the timeframe under consideration. We considered significant variables from the prior models. Lastly, we applied clinical judgment from experts and prior experience with the similar studies to finalize the predictor dataset [16]. We further divided the set into clinical, administrative, and psychosocial domain variables to test the predictivity of individual and combined domains. Our clinical dataset consisted of vital signs and certain laboratory values. The administrative dataset consisted of comorbidities and demographic variables that were captured routinely during hospitalization. The psychosocial dataset mainly consisted of psychiatric comorbidities including substance abuse and other variables indicating patient's behavioral and social situation. We gathered the data values for the variables that were closest to the date of discharge for the given episode. For the comorbidities, we individually tracked ICD-9-CM codes for the diseases and procedures that were suggested from the prior literature [4; 9; 12; 14]. The predictor dataset for each domain are shown in Tables 1, 2, and 3.

Statistical Analysis

The extraction of data from the EHR system yielded a sample size of 1210 admissions with 59 predictor variables after applying the exclusion criteria. We did not have any missing values for the dependent variable; a few independent variables such as BNP and blood glucose had up to 5% missing values. The missingness for the variables was tested and found to be missing at random. We used multiple imputation by chained equations resampled over five imputed datasets for the missing values [17]. We created separate dummy variables for missing values with higher rate of missingness for the categorical variables [13; 14]. This strategy allowed us to use all the available records in the analysis. Continuous variables were also examined for nonlinear effects and transformations were carried wherever necessary. We used the elastic net algorithm with combined least angle and ridge regression regularization for solving the logistic regression problem. The algorithm internally uses the shrinkage method for variable selection in a cross-validation scheme iterated with coordinate descent method [18]. The main advantage of this method is its usefulness in handling sparse data matrix with many variables. Thus, it allowed us to keep all the variables in our dataset. A 10-fold cross-validation scheme was used to derive and validate the models for the datasets in clinical, administrative, and psychosocial domains. All the results were reported for the validation dataset. All confidence intervals were calculated at 95% level using non-parameterized bootstrap adjustment. A bootstrapping process with 2000 iterations was used on the derivation cohort and the coefficients were averaged over this process. The discrimination using C-Statistics and predictive ability from lowest to highest deciles in the derivation and validation cohorts were calculated. We used R version 3.0.2 (The R Foundation for Statistical Computing Platform) for all analyses. The Veterans Affairs Palo Alto Health Care System's Institutional Review Board approved the research protocol.

Table 1 – Clinical Dataset

Clinical Predictors	Predictor Name	Total (n= 1210) Mean ± SD	Not Readm (n = 947)	Readm (n = 263)
Vitals	Heart Rate (beats/min)	80 ± 17	80 ± 17	81 ± 16
	Respiratory Rate (breath/min)	20 ± 3	20 ± 3	20 ± 3
	Systolic Pressure (mm of Hg)	127 ± 24	129 ± 24	120 ± 23
Labs	Plasma Glucose (mg/dL)	120 ± 42	120 ± 42	121 ± 43
	Serum Blood Urea Nitrogen (mg/dL)	33.9 ± 21.9	31.9 ± 20.2	41.0 ± 26.2
	Serum Creatinine (mg/dL)	1.6 ± 1.0	1.6 ± 1.0	1.7 ± 0.9
	Serum Sodium (mEq/L)	138 ± 3.6	138 ± 3.6	138 ± 3.5
	Serum Potassium (mEq/L)	4.1 ± 0.6	4.1 ± 0.5	4.1 ± 0.6
	Serum Albumin (g/dL)	3.0 ± 0.6	3.0 ± 0.6	3.0 ± 0.5
	Blood Hemoglobin (g/dL)	11.9 ± 2.1	11.9 ± 2.1	11.5 ± 1.9
	Blood Hematocrit (%)	35.6 ± 6.2	35.8 ± 6.4	34.7 ± 5.6
	Serum B-Natriuretic Peptide (pg/mL)	1027 ± 1180	1020 ± 1235	1053 ± 960

Results

White race (when documented) and male gender with an average age of 76.8 years predominantly characterized our cohort. The dataset represented 1210 patient admissions; 263 of which were readmitted indicating 21.74% (28.54% for repeated readmissions averaged over the study time) of 30-day readmission rate.

The lowest C-Statistics or Area Under Curve (AUC) (0.50, CI: [0.49, 0.51]) was obtained using predictors only from the psychosocial domain and the highest discrimination (0.84, CI: [0.83, 0.85]) was observed in the model with predictors combined from all the domains. Incremental increase in AUC was observed as (0.65, CI: [0.64, 0.66]), (0.71, CI: [0.69, 0.72]), and (0.77, CI: [0.75, 0.78]) for clinical, administrative,

and combined clinical and administrative predictors respectively. A graphical representation of AUC using Receiver Operating Curve (ROC) for various models for the validation cohort is shown in Figure 1. In our dataset, patients

who got readmitted were more likely to have prior HF diagnosis, more vascular and renal comorbidities, more discharges to skilled nursing facilities and had less outpatient visits than those who were not readmitted.

Table 2 – Administrative Dataset

Administrative Predictors	Predictor Name	Total (n=1210)	Not Readmitted (n = 947)	Readmitted (n = 263)
Demographic Characteristics	Sex:		n %	
	Female	1.82	20 1.65	2 0.17
	Male	98.18	927 76.61	261 21.57
	Age (years)	76.8 ± 12.4	Mean ± SD 76.9 ± 12.3	76.4 ± 12.6
	Race:		n %	
	White	32.31	300 24.79	91 7.52
	African American	4.79	34 2.81	24 1.98
	Hispanic	4.13	38 3.14	12 0.99
	American Indian & Alaska Native	0.25	2 0.17	1 0.08
	Asian	1.07	12 0.99	1 0.08
	Unknown	57.43	561 46.36	134 11.07
Payment Method	Medicare Insurance:		n %	
	No	49.92	462 38.18	142 11.74
	Yes	50.08	485 40.08	121 10.00
	Medi-Cal Insurance:		n %	
	No	97.86	932 77.03	252 20.83
	Yes	2.14	15 1.24	11 0.91
Comorbidities (Only Disease Groups Reported)	Prior Heart Failure:		n %	
	No	86.03	947 72.06	169 13.97
	Yes	7.77	0 0.00	94 7.77
	Diabetes Mellitus:		n %	
	No	57.18	548 45.28	144 11.90
	Yes	42.82	399 32.98	119 9.84
	Coronary Artery Disease:		n %	
	No	51.08	495 40.91	123 10.17
	Yes	48.92	452 37.36	140 11.56
	Ischemic Heart Disease:		n %	
	No	94.05	878 72.56	260 21.49
	Yes	5.95	69 5.70	3 0.25
	MI:		n %	
	No	93.80	1778 73.47	492 20.33
	Yes	6.20	116 4.79	34 1.41
	Valvular Heart Disease:		n %	
	No	90.62	1719 71.03	474 19.59
	Yes	9.38	175 7.23	52 2.15
	Cerebrovascular Disease/Stroke/TIA:		n %	
	No	98.46	3751 77.50	1039 21.46
	Yes	1.04	37 0.77	13 0.27
	Vascular/Circulatory Disease:		n %	
	No	87.47	4131 68.28	1161 19.19
	Yes	12.53	604 9.98	154 2.55
	Arrhythmias:		n %	
	No	53.48	524 43.31	123 10.17
	Yes	46.52	423 34.95	140 11.57
	Cardiomyopathy:		n %	
	No	82.97	802 66.28	202 16.69
	Yes	17.03	145 11.98	61 5.05
	Renal Disease or ESRD or Dialysis:		n %	
	No	77.77	1495 61.78	387 15.99
	Yes	22.23	399 16.49	139 5.74
	Chronic Lung Disease/COPD/Asthma:		n %	
	No	90.42	1707 70.54	481 19.88
	Yes	9.58	187 7.73	45 1.85
	Metastatic Cancer/Acute Leukemia/severe hematological disorder:		n %	
	No	93.94	2670 73.55	740 20.39
	Yes	6.06	171 4.71	49 1.35
	Liver Disease:		n %	
	No	97.19	918 75.87	258 21.32
	Yes	2.81	29 2.40	5 0.41

Table 3 – Psychosocial Dataset

Psychosocial Predictors	Predictor Name	Total (n=1210)	Not Readmitted (n = 947)	Readmitted (n = 263)
Social Characteristics	Marital Status:		n %	
	Married	37.85	359 29.67	99 8.18
	Divorced	31.32	291 21.05	88 7.27
	Widowed	16.28	155 12.81	42 3.47
	Never Married	10.99	106 8.76	27 2.23
	Separated	3.22	32 2.65	7 0.58
Behavioral Characteristics (with respect to Index Admission)	Unknown	0.34	4 0.34	0 0.00
	Discharge to Skilled Nursing Facility:		n %	
	No	94.21	916 75.70	224 18.51
	Yes	5.79	31 2.57	39 3.22
	Missed Clinic Visits in Prior Year:		n %	
	No	85.20	807 66.69	224 18.51
	Yes	14.80	140 11.58	39 3.22
	ED and O/P Visits in Prior Year:		n %	
	No	45.37	349 28.84	200 16.53
	Yes	54.63	598 49.42	63 5.21
	Admission in Previous Year:		n %	
	No	93.97	947 78.27	190 15.70
	Yes	6.03	0 0.00	73 6.03
	In-admission Telemetry Monitoring:		n %	
	No	69.59	662 54.71	180 14.88
	Yes	30.41	285 23.55	83 6.86
Comorbidities (Only Disease Groups Reported)	Mental Disorder(s):		n %	
	No	94.79	896 74.05	251 20.74
	Yes	5.21	51 4.22	12 0.99
	Alcohol Abuse:		n %	
	No	94.63	894 73.88	251 20.75
	Yes	5.37	53 4.38	12 0.99
	Drug Abuse:		n %	
	No	93.63	888 73.38	245 20.25
	Yes	6.37	59 4.88	18 1.49
	Protein Caloric Malnutrition:		n %	
	No	99.26	942 77.85	259 21.41
	Yes	0.74	5 0.41	4 0.33
	Functional Disabilities:		n %	
	No	99.51	942 77.85	262 21.66
	Yes	0.49	5 0.41	1 0.08

The final model stratified by predicted risk for 30-day readmission rate was calculated. Its plot shown in Figure 2 confirmed that the risk predictions were highly concordant across both the derivation and validation cohorts and thus confirmed absence of overfitting.

Discussion

This paper explores 30-day readmission risk model based on the already established idea from the PMI that a few diseases drive facility readmission rate and stratified risk assessment of patients would benefit from precise set of intervention guidelines for each stratum. Many models have been suggested on the basis of available data in patient records [9; 14] – this situation is made possible by the implementation of the Health Information Technology for Economic and Clinical Health (HITECH) act of 2009 in the United States that required EHR applications to incorporate many patient-centric data items—such as demographics, clinical data, and other psychosocial determinants of health—into the record.

Our results have also verified the claims of the other researchers who have combined claims data with clinical data that more predictors from different domains are preferable to consideration of fewer [14]. We confirmed the findings from the previously published studies [9; 11; 13-15] (AUCs: 0.69; NA; 0.60; 0.60; 0.61 respectively) by incrementally adding predictors from multiple domains and comparing the observed readmission rate and C-Statistics of our model to them.

Our work indicates that predictors combined from various domains have a strong influence on the prediction of 30-day readmission risk; at the other end of the spectrum, it also shows that predictors from a single domain tend to have limited impact on the risk prediction. These results also support the empirical observation that providers working in a multidisciplinary team in a patient-centered environment tend to reduce the risk of HF readmissions.

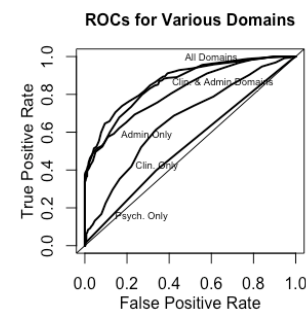


Figure 1 – Validation Cohort ROCs

Our study has some limitations. First, we were limited by the demographics of our cohort. Second, our study was carried out at one health system although all the previously published model results were also single center studies. As the efforts of building large-scale clinical research data networks with

multi-facility collaborations come to fruition in near future, we will be able to collectively overcome this limitation. Third, only structured predictors were considered in our study and future studies might benefit by exploring valuable information available in other components—such as imaging reports and clinical notes—of the EHR system. Lastly, 30-day readmissions that might have happened in other facilities were not considered in our analysis, although the proportion of such readmissions in the overall cohort was deemed small based on empirical observations.

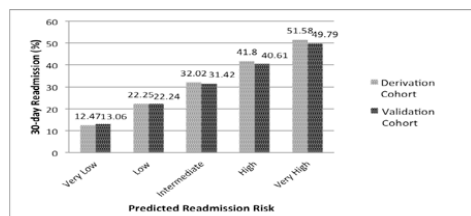


Figure 2 – Final Model Validation Cohort Performance

Conclusion

Risk of frequent readmissions in HF population is an important issue for patients, clinicians, and hospital administrators. Our study has shown that the use of information from multiple domains in the hospital EHR system helps to build a risk model for precise interventions for HF readmission, is effective, and applies novel statistical techniques to this domain. Additional risk factors using unstructured data could be incorporated into our model in the future to strengthen the foundation of a PMI effort started here.

Acknowledgements

This project was carried out under the joint IRB approved by VA Palo Alto Health Care System (VAPAHCS) and Stanford University (eProtocol# 12015).

References

- [1] NIH and Committee, NIH-wide Strategic Plan Fiscal Years 2016-2020, in, Bethesda, MD, USA, 2015.
- [2] N. Fida and I. L. Pina, Trends in heart failure hospitalizations, *Curr Heart Fail Rep* **9** (2012), 346-353.
- [3] P. A. Heidenreich, J. G. Trogon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, D. M. Lloyd-Jones, S. A. Nelson, G. Nichol, D. Orenstein, P. W. Wilson, Y. J. Woo, Committee American Heart Association Advocacy Coordinating, Council Stroke, Radiology Council on Cardiovascular, Intervention, Cardiology Council on Clinical, Epidemiology Council on, Prevention, Arteriosclerosis Council on, Thrombosis, Biology Vascular, Cardiopulmonary Council on, Care Critical, Perioperative, Resuscitation, Nursing Council on Cardiovascular, Disease Council on the Kidney in Cardiovascular, Surgery Council on Cardiovascular, Anesthesia, Care Interdisciplinary Council on Quality of, and Research Outcomes, Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association, *Circulation* **123** (2011), 933-944.
- [4] C. M. O'Connor, W. T. Abraham, N. M. Albert, R. Clare, W. Gattis Stough, M. Gheorghade, B. H. Greenberg, C. W. Yancy, J. B. Young, and G. C. Fonarow, Predictors of mortality after discharge in patients hospitalized with heart failure: an analysis from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF), *Am Heart J* **156** (2008), 662-673.
- [5] G. M. Felker, J. D. Leimberger, R. M. Califf, M. S. Cuffe, B. M. Massie, K. F. Adams, M. Gheorghade, and C. M. O'Connor, Risk stratification

after hospitalization for decompensated heart failure, *Journal of Cardiac Failure* **10** (2004), 460-466.

- [6] J. Zhang, K. M. Goode, A. Rigby, A. H. Balk, and J. G. Cleland, Identifying patients at risk of death or hospitalisation due to worsening heart failure using decision tree analysis: evidence from the Trans-European Network-Home-Care Management System (TEN-HMS) study, *Int J Cardiol* **163** (2013), 149-156.
- [7] C. M. O'Connor, R. J. Mentz, G. Cotter, M. Metra, J. G. Cleland, B. A. Davison, M. M. Givertz, G. A. Mansoor, P. Ponikowski, J. R. Teerlink, A. A. Voors, M. Fiuzat, D. Wojdyla, K. Chiswell, and B. M. Massie, The PROTECT in-hospital risk model: 7-day outcome in patients hospitalized with acute heart failure and renal dysfunction, *Eur J Heart Fail* **14** (2012), 605-612.
- [8] L. M. Yamokoski, V. Hasselblad, D. K. Moser, C. Binanay, G. A. Conway, J. M. Glotzer, K. A. Hartman, L. W. Stevenson, and C. V. Leier, Prediction of rehospitalization and death in severe heart failure by physicians and nurses of the ESCAPE trial, *J Card Fail* **13** (2007), 8-13.
- [9] P. S. Keenan, S. L. Normand, Z. Lin, E. E. Drye, K. R. Bhat, J. S. Ross, J. D. Schuur, B. D. Stauffer, S. M. Bernheim, A. J. Epstein, Y. Wang, J. Herrin, J. Chen, J. J. Federer, J. A. Mattera, Y. Wang, and H. M. Krumholz, An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure, *Circ Cardiovasc Qual Outcomes* **1** (2008), 29-37.
- [10] H. M. Krumholz, Y. W. Chen, Yun, Viola Vaccarino, Martha J. Radford, and Ralph I. Horwitz, Predictors of readmission among elderly survivors of admission with heart failure, *American Heart Journal* **139** (2000), 72-77.
- [11] H. M. Krumholz, L. Horwitz, C. Partovian, Z. Lin, J. Herrin, J. Grady, M. Conover, J. Montague, C. Dillaway, K. Bartczak, J. Ross, S. Bernheim, and E. Drye, Hospital-Wide (All-Condition) 30-Day Risk-Standardized Readmission Measure, in, Yale New Haven Health Services Corporation/Center, New Haven, CT, 2008.
- [12] E. F. Philbin and T. G. DiSalvo, Prediction of hospital readmission for heart failure: Development of a simple risk score based on administrative data, *JACC* **33** (1999), 1560-1566.
- [13] R. Amarasingham, B. J. Moore, Y. P. Tabak, M. H. Drazner, C. A. Clark, S. Zhang, W. G. Reed, T. S. Swanson, Y. Ma, and E. A. Halm, An Automated Model to Identify Heart Failure Patients at Risk for 30-Day Readmission or Death Using Electronic Medical Record Data, *Med Care* **48** (2010), 981-988.
- [14] B. G. Hammill, L. H. Curtis, G. C. Fonarow, P. A. Heidenreich, C. W. Yancy, E. D. Peterson, and A. F. Hernandez, Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization, *Circ Cardiovasc Qual Outcomes* **4** (2011), 60-67.
- [15] M. H. Chin and L. Goldman, Correlates of early hospital readmission or death in patients with congestive heart failure, *Am J Cardiol* **79** (1997), 1640-1644.
- [16] S. M. Mahajan, P. Burman, and M. A. Hogarth, Analyzing 30-Day Readmission Rate for Heart Failure Using Different Predictive Models, *Studies in Health Technology and Informatics* **225** (2016), 143-147.
- [17] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, Multiple imputation by chained equations: what is it and how does it work?, *Int J Methods Psychiatr Res* **20** (2011), 40-49.
- [18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2008.

Address for correspondence

Satish M. Mahajan, Betty Irene Moore School of Nursing, University of California Davis, Education Building, 4610 X Street, Sacramento, CA 95817. Email: smmahajan@ucdavis.edu

CrowdMapping: A Crowdsourcing-Based Terminology Mapping Method for Medical Data Standardization

Huajian Mao, Chenyang Chi, Boyu Huang, Haibin Meng,
Jinghui Yu, Dongsheng Zhao

Institute of Health Service and Medical Informatics, Academy of Military Medical Sciences of Chinese PLA, Beijing, China

Abstract

Standardized terminology is the prerequisite of data exchange in analysis of clinical processes. However, data from different electronic health record systems are based on idiosyncratic terminology systems, especially when the data is from different hospitals and healthcare organizations. Terminology standardization is necessary for the medical data analysis. We propose a crowdsourcing-based terminology mapping method, CrowdMapping, to standardize the terminology in medical data. CrowdMapping uses a confidential model to determine how terminologies are mapped to a standard system, like ICD-10. The model uses mappings from different health care organizations and evaluates the diversity of the mapping to determine a more sophisticated mapping rule. Further, the CrowdMapping model enables users to rate the mapping result and interact with the model evaluation. CrowdMapping is a work-in-progress system, we present initial results mapping terminologies.

Keywords:

Crowdsourcing; Medical Informatics; Vocabulary, Controlled

Introduction

Analyzing clinical data across medical centers requires mapping local and idiosyncratic information into standardized terminologies. Without this step, clinical data cannot be exchanged, shared, integrated or used in a meaningful way. There are lots of standard terminology systems for representing clinical data, including Logical Observation Identifiers Names and Codes (LOINC) [1], Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) [2] and others, which are now used as international standards. However, in lots of the healthcare organizations, they still use local and idiosyncratic dictionaries, which always reduces the ability to integrate data from multiple organizations. Therefore, in order to understand generated medical information in digital systems, healthcare organizations must often translate the local data into standardized terminologies. The process of translating the local data into standardized terminologies is called mapping. A mapping represents a set of terminology mediation strategies used by clinical and public health organizations to enable health information exchange (HIE) within and among health enterprises [3].

At the Institution of Health Service and Medical Informatics of Academy of Military Medical Sciences (AMMS), we are undertaking this mapping process because our medical data center is involved in the creation of a nation-wide health information exchange that requires sharing clinical data across more than 200 first-class hospitals at Grade 3 in China. In the hospital information systems of these 200 hospitals, they use

hundreds of dictionaries (e.g., diagnosis, operation, and other terminology dictionaries). Meanwhile, a large part of the terminologies' dictionaries contain a huge amount of entities. How to standardize the terminologies from these 200 hospitals are the most challenging work when integrating their data.

Although there are deviations in terminology standards for different hospitals, there are a number of terms that are generally accepted. In this way, although different local terminology systems are used in different hospitals, a large part of the terminologies are the same. There would be a lot of duplicate work on terminology mapping if each of these 200 hospitals respectively standardizes its terminology system into a specified one.

Besides, if the mapping work is done separately by each of the hospitals, it would probably generate different mappings for some selected terminology. For example, hospital A might map LV_a to ST_a terminology in the standard and hospital B might map LV_a to a different ST_b terminology. How to eliminate the error mapping for the 200 hospitals and ensure that the local vocabularies in the different hospitals are correctly mapped to the same standard terminologies is challenging work. It would significantly improve the precision of mapping if we had some mechanism to verify the mapping and revise the error mapped ones.

Meanwhile, as there are a huge number of terminology entities in the mappings, it would be also a huge task to map all of them to standard terminologies. However, in practice, only a small part of the local items would be used in the actual healthcare systems. It would improve the efficiency if we considered mapping the terminologies based on their usage frequency.

In this paper, we propose a crowdsourcing based terminology mapping method, CrowdMapping, to help healthcare providers map their local and idiosyncratic data to standard terminologies. This would provide a better use of available medical data for analysis, integration and so on. First, CrowdMapping allows users from different organizations to specify mapping candidates in a standard terminology system like ICD-10. The model uses the mappings from the users and evaluates the diversity of the mapping to determine a most-selected, final best mapping rule if the diversity is less than a threshold value. Otherwise, if the diversity is too large, then some new mappings for the local terminology are generated from some more sophisticated experts. Further, the CrowdMapping model also enables users to give negative ratings to the final selected mapping rule if that mapping would be likely be wrong. CrowdMapping will recalculate the score for that terminology by subtracting a value which is related to the user's reputation. Then, the final best mapping rule is recalculated. CrowdMapping is a work-in-progress system. It will be made public when the system is stable enough for usage.

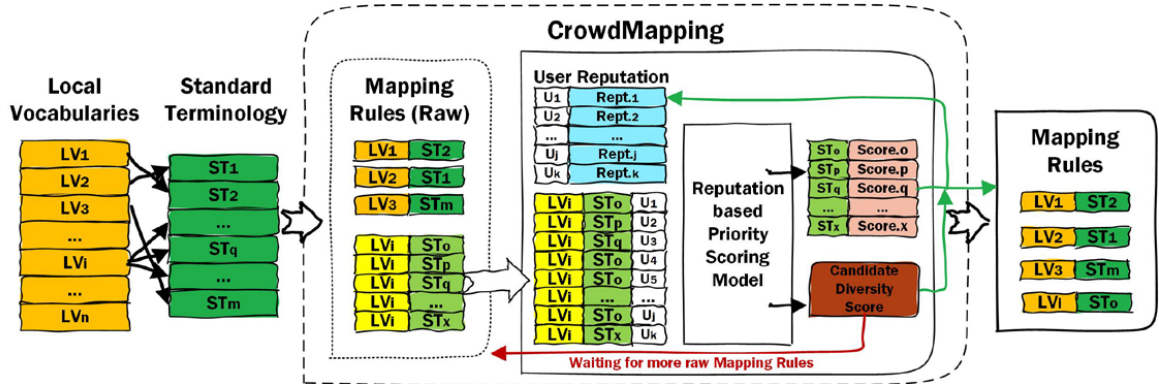


Figure 1 – CrowdMapping Architecture

Methods

Traditionally, standardizing local vocabularies is done by assigning the most likely terminology in the standard system by users. For example, as shown in Figure 1, a user can map LV_1 in the local vocabulary to ST_2 in the standard terminology, same with LV_2 and LV_3 . However, mapping the local vocabulary to a standard terminology is not a trivial work. For example, different users might map some local vocabulary (LV_i in the figure) to different standard terminologies (for example, mapping LV_i to $ST_o, ST_p, ST_q, \dots, ST_x$ and so on). How to select the most likely candidate as the standard terminology for LV_i would become a problem in this scenario. CrowdMapping proposes a crowdsourcing based solution to solve this problem.

Overview

The target of CrowdMapping is to provide healthcare providers with a platform to map their local vocabularies, stored in dictionaries, to standard terminologies with high accuracy while keeping the mapping process simple and maintaining high efficiency. There are two basic design ideas in CrowdMapping.

1. CrowdMapping allows different experts to map a local vocabulary to several different candidates. It then uses a rating model to calculate a diversity value for the candidates. If the diversity value is less than a given threshold, a top-rated candidate will be automatically selected as the final terminology for the local vocabulary.
2. The volume of local data to be mapped is always very large, but fortunately, not all of the vocabularies in the local vocabulary systems are used; the usage of the local vocabularies are always biased. Mapping a small part of the local vocabularies would likely standardize most of the local data. With this evidence, CrowdMapping determines the next mapping vocabulary with terminology usage frequency in the local dataset.

Figure 1 shows the design of CrowdMapping. CrowdMapping system consists of three major parts, which are local vocabularies and standard terminology system repositories; a user reputation based priority scoring model; and a final mapping rule repository. The local vocabulary repository contains all the vocabularies which are local and are going to be mapped, while the standard terminology repository is the target to map the local ones to. The standard terminology systems are usually standard terminology systems like ICD-10, SNOMED and so on.

The users map the local vocabulary to standard terminology with their effort to make their decision of finding the most likely one in their opinion for the local one. CrowdMapping then records the user decisions (the raw mapping rule with the tag of which user made the mapping rule in the figure), and passes the raw user mapping rules to a reputation based priority scoring model with the user reputation values. The scoring model will calculate a candidate diversity score for each of local vocabularies (like LV_i). Besides, an ordered candidates queue, the priority scores are calculated by the scoring model. After the calculation, if the diversity is smaller than a given threshold a top-priority standard terminology will be selected as the final mapping value for the local vocabulary. For example, as shown in the figure, if the diversity is small, and ST_o has the highest score, then ST_o will be selected as the mapping value of LV_i . However, when the diversity value is larger than the given threshold, the decisions from different users are not convergent signifying different opinions for the mapping of that local vocabulary. So more raw mapping rules are needed to make the final mapping decision. CrowdMapping will recalculate the diversity and priority scores when a new mapping rule is made for that local vocabulary until the diversity is small enough. After this progress, CrowdMapping will generate a final mapping rules repository for user.

Reputation Based Priority Scoring

In order to describe the reputation based priority scoring model, we would like to define several terms.

User Reputation

User reputation is a value describing the expert level of a user on mapping the vocabularies. If a user always maps local vocabularies to correct standard terminologies, then he will get a higher reputation. Suppose user A has worked on the CrowdMapping platform for some time, and he/she has mapped a set of local terminologies to the standard terminologies. Let's suppose the mapping history of user A is:

```
val mappings = Set{ $m_1, m_2, \dots, m_n$ }
```

In which, user A has mapped k local vocabularies to standard terminologies correctly, and $n - k$ to error ones. Besides, as we will discuss in the *Mapping Rule Revision* section, the mapping rules can be '+1' or '-1'. Let's suppose the support number of mapping m_i is p_i , and the negative support number of mapping m_i is n_i . Then, if the mapping m_i is selected as a correct mapping, all the p_i users who have '+1'ed this mapping will be processed as having mapped correctly one more time, while the n_i error mapped mapping number of the '-1'ed users for the mapping m_i will be increased.

Now let's suppose that user A has mapped k local vocabularies correctly, and $n - k$ wrong, and has '+1' ed pc correctly, and pw wrong, and has '-1' ed nc correctly and nw wrong. Then we define the reputation r_i of user A as:

$$r_i = \alpha \frac{k}{k + n - k} + \beta \frac{pc}{pc + pw} + (1 - \alpha - \beta) \frac{nc}{nc + nw}$$

where α and β are two parameters which are configured by the CrowdMapping system.

Mapping Set Diversity

Mapping set diversity describes a status of the current mappings for a specific vocabulary. If there are many different mappings for a local vocabulary, that means users having different opinions on the mappings for that local vocabulary will have a high diversity. Otherwise, if all the mappings from different users are the same, then the diversity of the mapping set is 0. Diversity has been used in various areas [4-6]. In the system implementation, Simpson's Diversity Index (SDI) [4] is used to calculate the diversity of the mapping set.

Suppose that the local vocabulary LV_i is mapped to k potential standard terminologies which are $\{ST_i^1, ST_i^2, \dots, ST_i^k\}$, and have n_i^1 users mapping LV_i to ST_i^1 , n_i^2 users mapping LV_i to ST_i^2 , and n_i^k users mapping LV_i to ST_i^k . Then according to the algorithm of Simpson's Diversity Index, we get

$$D_i = 1 - \frac{n_i^1(n_i^1 - 1) + n_i^2(n_i^2 - 1) + \dots + n_i^k(n_i^k - 1)}{N(N - 1)}$$

where $N = n_i^1 + n_i^2 + \dots + n_i^k$.

With this algorithm, the diversity value of the mapping set for LV_i is mapped into a range of [0, 1]. For example, if all the users map the LV_i to ST_i^1 , then $N = n_i^1$, so $D_i = 1 - \frac{n_i^1(n_i^1 - 1)}{N(N - 1)} = 0$. Or, if all the users map the LV_i to different standard terminologies, then all the n_i^j s equals to 1, which will lead to

$$D_i = 1 - \frac{\sum n_i^j(1 - 1)}{N(N - 1)} = 1 - 0 = 1.$$

Mapping Confidence

Mapping confidence defines the level of how much of the mapping can be trusted. Mapping confidence is always calculated on the selected final mapping rule which has the highest score. It is affected by the mapping count, user revision score, and the mapping diversity. A higher mapping confidence means a higher probability that the mapping can be trusted.

Let's continue with the example used in Mapping Set Diversity. Suppose the mapping M_i^j from LV_i to ST_i^j is the mapping which has the highest score. Suppose that the count of the users who map LV_i to ST_i^j is c_i^j , and the diversity of LV_i is D_i , and pc_i^j users '+1' the mapping M_i^j , and nc_i^j users '-1' the mapping M_i^j . The confidence of the mapping M_i^j can be calculated as:

$$C_i = \frac{pc_i^j}{pc_i^j + nc_i^j} * D_i * \frac{count_i^j}{count_{MIN}}$$

Only the mapping with a high confidence larger than a threshold C_{MIN} will be selected as a final mapping for the local vocabulary and be put into the mapping rule repository.

Score Calculation

Based on the definitions above, suppose that n users (U_1, U_2, \dots, U_n) are working on mapping the local vocabularies LV_i to standard terminologies. Let's annotate the reputation of U_i as r_i . The user-generated k mappings for LV_i , $\{ST_i^1, ST_i^2, \dots, ST_i^k\}$ (where $k \leq n$). Suppose that those who have mapped the LV_i to ST_i^j are the users (U_a, U_b, \dots, U_x). Then we will calculate a

temporal reputation summary of the mapping M_i^j from LV_i to ST_i^j :

$$Reput_j = \sum_{i=a}^x r_i$$

and then the score for the mapping M_i^j is calculated as:

$$Score_j = \frac{Reput_j}{\sum_{i=1}^k Reput_i}$$

The reputation based priority scoring model will calculate all the mappings and their scores and generate an scored terminology map as $\{ST_o:Score_o, ST_p:Score_p, ST_q:Score_q, \dots, ST_x:Score_x\}$.

After the scores are computed, the CrowdMapping method will calculate the mapping confidence value for the mapping with the highest score. CrowdMapping will compare it with a given threshold value, if the diversity score is less than the threshold, CrowdMapping will select the standard terminology ST_o with the highest score (for example, $Score_o$) as the mapping to terminology for the local vocabulary (for example, LV_i in the figure). CrowdMapping will also update the user reputation table according to the last mapping decisions. In this way, the users who have mapped LV_i to ST_o will get a higher reputation value, while the users mapped LV_i to other standard terminology will be reduced on their reputation. However, if the confidence value of the mapping is too small, suggesting that the users do not have a consistent decision on how to map LV_i to the standard terminology, CrowdMapping will not do the following steps as described above. It will wait for more user raw mapping rules.

Algorithm 1 shows the process of reputation based priority scoring method of finding the most likely mapping M_i^j for LV_i . The CrowdMapping system will use this algorithm to process the mappings for all the local vocabularies.

With these processes, a final mapping decision will be made based on all the user mapping decisions, as shown in the Mapping Rules in the figure.

Algorithm 1: reputation based score algorithm for LV_i

mappings \leftarrow all the mappings of LV_i with user info

Di \leftarrow calculate diversity of the *mappings*

if *Di* > *Dthreshold*, then

users \leftarrow get users of the *mappings*

 for each user *in users*, do

ri \leftarrow calculate user reputation for each user

 end

scoredMap \leftarrow init scored terminology mappings

 for each mapping *in mappings*, do

Scorei \leftarrow calculate *Score* of mapping

Ci \leftarrow calculate *confidence* of mapping

 if *Ci* > *CMIN*, then

 Add (*STi*, *Scorei*) to *scoredMap*

 end

 end

orderedScoredMap \leftarrow sort *scoredMap* by score

 return top of *orderedScoredMap*

else

 return NULL

end

Frequency Based Local Vocabulary Candidate

Local vocabularies are always having a large volume of entries. In practice, only a limited part of the local vocabularies are used in the healthcare systems. So different entries in the vocabularies have different significance for health care systems. It would significantly improve the work efficiency if the most 'important' entries in the local vocabulary are mapped first. So CrowdMapping defines a usage frequency based priority value for local entry candidate selection. If an entry in the vocabulary is used quite frequently in a health care setting it will have a higher priority for mapping.

However, the frequency is calculated from online health care systems. In some case, users might not be able to get the frequency. In these situations CrowdMapping will use a similar frequency statistic from another health care organization. If no organization has the frequency information, then equal priority will be assigned to the local vocabulary entries. In this way, the next local vocabulary entry to be processed is selected randomly.

Mapping Rule Revision

After the steps described above are performed a usable mapping rules repository will be generated. And users from different healthcare organizations can use these mapping rules to automatically standardize their local vocabularies. However, CrowdMapping is a machine-based automation tool, so there might exist erroneous mappings in the system-generated rules. An official selected mapping may also have expired and/or become invalid. For these reasons, CrowdMapping has a mapping rule revision module. In this module, the users can '+1' a mapping rule or '-1' it. And these actions will trigger the mapping confidence computation. If there are too many users '-1' ing that mapping rule, the mapping confidence will decrease. When the confidence is lower than a threshold, the mapping rule will be removed from the final mapping rule repository to the raw mapping rule repository to await more user mapping decision.

Results

As CrowdMapping is still a work-in-progress project, we have not performed large scale testing. However, with using the key design idea in a demo platform, we found that CrowdMapping combines the mapping decisions from multiple users and selects the most likely rule for the local vocabularies which makes the rule mapping decision more precise.

When the platform is ready, we will evaluate CrowdMapping with the traditional mapping method on mapping precision and the processing time. Also, we will make CrowdMapping as a public web application. All the users from healthcare organizations can use our service together with the effort from the experts all over the world.

Discussion

Terminology mapping is important work for the medical informatics community. There is already a large body of work describing various aspects of mapping local codes to standard terminology systems from both academic and industry researchers.

Several studies [7-11] have evaluated different automated tools to assist with mapping local vocabularies to other terminology systems. Yet, even with the best available automated tools, expert human review is still needed to resolve computer generated candidate mappings. Also, because local and standard vocabularies evolve, the burden of maintaining the

mappings is significant, ongoing, and easily underestimated. Therefore, all healthcare organizations whether data senders, receivers, or both, require people, processes, and tools to support mapping activities.

It is a complex and resource intensive job to map local terms to standard terminology systems. Even a sophisticated person with a good understanding of the corresponding terminology system, he/she might lack the specific knowledge required to correctly map all of their local terminologies [3,12,13]. How to efficiently combine the effort from the experts of the institutions to improve the mapping precision is a necessary work. Crowdsourcing has been used in many fields to improve the work efficiency. In the terminology mapping, several work [3,12,13] have been done to make mapping more efficient and effective with crowdsourcing. However, they are strongly integrated with special standardized terminologies, for example, LOINC.

CrowdMapping allows users from different organizations to specify mapping candidates in standard terminology systems like ICD-10. CrowdMapping uses the mappings from the users and evaluates the diversity of the mapping to determine a most-frequently selected final best mapping rule. This strategy would likely improve the accuracy of the mapping. Mapping set diversity and the mapping confidence model are used in CrowdMapping to automatically divide the 'hard' mapping tasks from the 'easy' ones (a local entry with a high mapping diversity is always a 'hard' mapping task). CrowdMapping can let the 'right' person to do the 'right' thing; the 'hard' mapping tasks are assigned to sophisticated experts to make the mapping more precise.

Conclusion

In this paper, we propose CrowdMapping, a platform for generating terminology mapping rules that leverages the crowd effort. CrowdMapping uses an algorithm which considers both the user reputation and crowd selection, which likely leads to a more confident mapping result. Using the demo program, we found that CrowdMapping reduces the mapping time and increases precision. Terminology mapping is a very common step for medical data analysis, we will deploy an online service for public use as soon as the platform is more stable.

Acknowledgements

The authors would like to thank the anonymous reviewers for their insight, comments, and kindsuggestions. Their valuable comments helped improve this paper a lot.

References

- [1] R.A. Cote, College of American Pathologists, *Systematized Nomenclature of Medicine*, College of American Pathologists, 1977.
- [2] B.E. Dixon, J. Hook, D.J. Veerman, Learning from the crowd in terminology mapping: the loinc experience, *Lab Med* **46** (2015), 168-174.
- [3] C. McDonald, S. Huff, J. Suico, K. Mercer, Logical observation identifiers names and codes (loinc R) users' guide, in: Regenstrief Institute, Indianapolis, 2004.
- [4] E.H. Simpson, Measurement of Diversity, *Nature* **163** (1949), 688.
- [5] M. Nei, Analysis of gene diversity in subdivided populations, in: *Proceedings of the National Academy of Sciences*, 1973, pp. 3321-3323.
- [6] A.E. Magurran, *Why diversity? In ecological diversity and its measurement*, 1988.
- [7] A.N. Khan, S.P. Griffith, C. Moore, D. Russell, A.C. Rosario, J. Bertolli, Standardizing laboratory data by mapping to loinc., *J Am Med Inform Assoc* **13** (2006), 353-355.
- [8] L.M. Lau, K. Johnson, K. Monson, S.H. Lam, S.M. Huff, A method for the automated mapping of laboratory results to loinc., in: *Proceedings of the AMIA Symposium*, 2000.
- [9] J.Y. Sun, Y. Sun, A system for automated lexical mapping., *J Am Med*

Inform Assoc **7** (2006), 334-343.

- [10] K.A. Zollo, S.M. Huff, Automated mapping of observation codes using extensional definitions, *J Am Med Inform Assoc* **7** (2000), 586-592.
- [11] C. Zunner, T. Burkle, H.U. Prokosch, T. Ganslandt, Mapping local laboratory interface terms to loinc at a german university hospital using relma v. 5: a semi-automated approach, *J Am Med Inform Assoc* **20** (2013), 293-297.
- [12] M.C. Lin, D.J. Vreeman, C. McDonald, S.M. Huff, A characterization of local loinc mapping for laboratory tests in three large institutions, *Method Inform Med* **50** (2011), 105.
- [13] D.M. Baorto, J.J. Cimino, C.A. Parvin, M.G. Kahn, Combining laboratory data sets from multiple institutions using the logical observation identifier names and codes (LOINC), *Int J Med Inform* **51** (1998), 29-37.

Address for correspondence

Dongsheng Zhao, Mail: dszhao@bmi.ac.cn

Case Report Form Reporter: A Key Component for the Integration of Electronic Medical Records and the Electronic Data Capture System

Yasushi Matsumura^{a,b}, Atsushi Hattori^b, Shiro Manabe^{a,b}, Daiyo Takahashi^b,
Yuichiro Yamamoto^{a,b}, Taizo Murata^c, Akito Nakagawa^a, Naoki Mihara^a, Toshihiro Takeda^a

^a Medical Informatics, Osaka University Graduate School of Medicine, Osaka, Japan

^b MKS Ltd, Osaka, Japan

^c Division of Medical Informatics, Osaka University Hospital, Osaka, Japan

Abstract

To improve the efficiency of clinical research, we developed a system to integrate electronic medical records (EMRs) and the electronic data capture system (EDC). EDC is divided into case report form (CRF) reporter and CDMS with CRF receiver with data communication using the operational data model (ODM). The CRF reporter is incorporated into the EMR to share data with the EMR. In the data transcription type, doctors enter data using a progress note template, which are transmitted to the reporter template. It then generates the ODM. In the direct record type, reporter templates open from the progress note and generate narrative text to make record in the progress note. The configuration files for a study are delivered from the contents server to minimize the setup. This system has been used for 15 clinical studies including 3 clinical trials. This system can save labor and financial costs in clinical research.

Keywords:

Electronic Health Records; Clinical Study; Data Systems

Introduction

We need a method for transcribing the data recorded in a patient's medical record to a predetermined form for submission to a party outside of the hospital. In the era of managing medical records via a paper medical chart, the data written in the record needed to be transcribed by hand to another piece of paper; however, in the era of electronic medical records (EMRs), we now need a way to fill out forms without redundant data input. This function is particularly important in clinical research and would also be useful when sending a report to external organizations when for example following up a patient in collaboration with other medical facilities.

In clinical research, such as clinical trials, observational studies and registries, data for pre-determined items are systematically collected for patients who meet the entry criteria. In clinical trials, source data verification (SDV) by a monitor is required to confirm the consistency of the data in the case report form (CRF) with those in the medical record. This increases the costs associated with a study. In observational studies and registries, data are often entered without the support of clinical research coordinators (CRCs); as such, redundant entry in the medical record and electronic data capture system (EDC) is a major obstacle to promoting these studies. As registries become more popular, we may encounter instances where one case needs to be submitted to multiple registries, which results in redundant data input.

Interoperability between medical care and clinical research would be advantage, but has not been achieved in a wide scale

[1]. The Clinical Data Interchange Standard Consortium (CDISC) Electronic Source Data Interchange Group identified issues that may inhibit adoption, explored the values and benefits of implementing standards for data acquisition, exchange and archive of eSource [2]. We developed and already reported on a system to achieve these issues [3]. In our model, the EDC is divided into a CRF reporter, which functions as a data input and transmission unit, and a clinical data management system (CDMS) with CRF receiver, which functions as a data reception and storage unit. Data communication between the CRF reporter and the CRF receiver is achieved via an operational data model (ODM) developed by CDISC [4]. The CRF reporter is incorporated into an EMR system that can automatically retrieve the data recorded in the EMR and output narrative text generated from the data included in the CRF reporter to the medical record. In order to automate setup, a template master, an ODM and a study configuration file are distributed via a network from a content server to the hospitals participating in the study (Figure 1). Since the previous report, we have added several functions and refined the system further. In this paper, we describe our system, with focus on the CRF reporter, which is a key component for integration of EMRs and EDC.

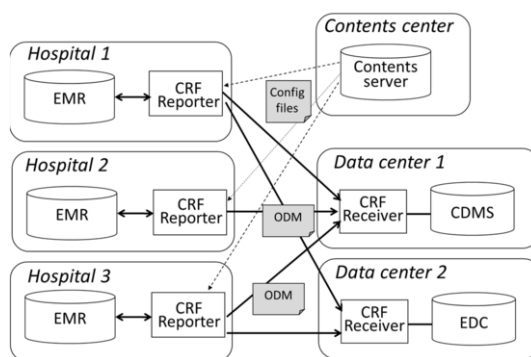


Figure 1 – Components of the System.

Methods

System Overview

There are two types of CRF reporters: a data transcription type, and a direct record type. The functions of each are outlined below.

Data Transcription Type

When recording progress notes, the data used for clinical research are input using a template. The CRF reporter is an application independent of the EMR. In the CRF reporter, after selecting a study and a subject, an event list is displayed. One event consists of several forms. When a form is selected, the corresponding template is displayed, along with a data retrieval button. By clicking this button, the data entered by the template for the progress note or data (e.g. laboratory test data) stored in the EMR database are retrieved and used to populate the items in the template. When the input of the template set corresponding to a study event is completed, the input data are mapped to the ODM, which is sent to the data center (Figure 2).

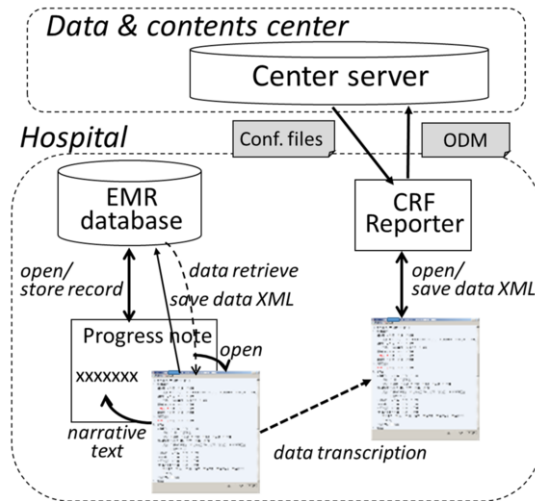


Figure 2 – Configuration of Data Transcription Type.

Direct Record Type

The start button of the CRF reporter is placed in the progress note. Clicking this button brings up a list of studies in which the patient and the user are participating. After selecting a study, a list of events with forms is opened. Selecting a form displays the corresponding template. After data input is completed, the narrative text is generated and added to a frame in the progress note. The subsequent process is the same as that of the data transcription type (Figure 3).

Input Template Function

The input template is a key module for this system. We developed “dynamic template” to acquire medical records in structured form [5]. If any abnormalities are noted in the observation items, a detailed description of the abnormality is required; this allows the medical description to be presented in a hierarchical tree structure. Therefore a template where the next input item changes according to the selected value is suitable for medical description.

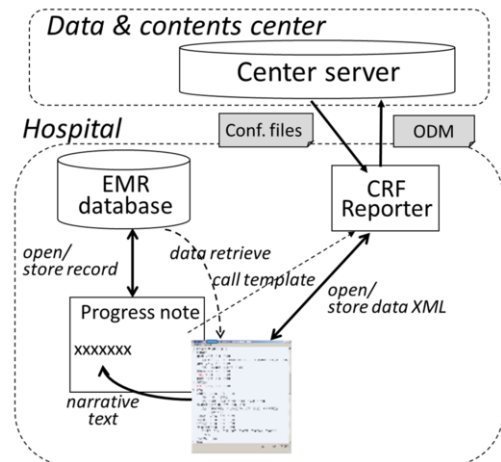


Figure 3 – Configuration of Direct Record Type.

To make it easier for the user to read the data entered by the template, natural narrative text is generated from the input data using natural language generation technology. The template outputs a medical record displayed with descriptive expressions, which is easier for the user to read. It also outputs the data as an XML file (data XML). When modifying the data, the input state is reproduced using the data XML.

The content of the template is controlled by the template master, which is also saved in the XML file (template XML). Therefore, when a different template XML is input into the template engine, different templates are displayed.

Data Retrieval Interface

In the EMR database, data such as patient characteristics (height, weight, etc.) and laboratory test results are stored. These data frequently need to be collected into templates. When retrieving laboratory test results or similar, an assigned test item code is required in order to obtain the value for a designated test item. However, the test item code differs among hospitals. Therefore, a conversion table (correspondence table between a universal keyword and the local code) is created at each hospital. This table also has a multiplied number to convert the value in local units to standard units and has a correspondence table to convert local enumerated values to standard ones.

In the template master, the parameters for data retrieval, such as the test item code, search period, and selection method, are described. In the template master, the test item code is described in the notation generally conforming to the CDISC-controlled terminology. The template engine obtains the local code for the test item and the information for converting the value into the standard form from the conversion table. Then it passes the parameters including the local code, search period, and selection method to the interface module. The selection methods are for example selecting one latest value, one oldest value or user's selection. If there are several corresponding data in a given search period, the data picker module presents them all to the user for selection.

The template can also retrieve values entered in other templates in the past by designating the element code with the template code of the source template or the concept code if it is written to the corresponding element.

Data Mapping to the eCRF

In our system, the ODM was used as the eCRF form outputted from the CRF reporter. After the completion of input, the template engine generates a dataset in a tree structure, where each leaf corresponds to a value and each node corresponds to an item. The value of a node is the narrative text generated from the data under the node. For each ItemData of the ODM, the corresponding value of the leaf or node can be mapped. For example, when a value of “the date of occurrence” is entered by selecting the year and month from a combo box on the template, the leaf value is year = 2010, month = 10, and the node value is “October 2010”. Thus, if the ItemData value corresponding to “the date of occurrence” is requested in a string, “October 2010” can be mapped to it.

ODM Transmission

First, we must enter subjects who meet the inclusion and exclusion criteria. At this point, the subject key is determined on the hospital side or issued on the server side; in our system, both are possible. The CRF reporter performs ODM transmission to the CDMS/EDC in the data center. To check the contents of the ODM, a readable case report form is also generated. After checking the contents, the ODM is transmitted to the center server.

Originally, the CRF reporter sent the ODM to one CDMS; however, it must now be able to send the ODM to any CDMS/EDC equipped with an ODM import function in datacenters, depending on the study. Because the communication procedures of ODMs differ, CRF reporters need to be able to switch transmission modules according to the destination center server.

User Management

The CRF reporter needs to manage local users as well as center server users linked to local users. When logging into the CRF reporter, a local user table is checked, but when creating and sending the ODM, the center server user is used instead. The user ID for the center server is recorded in the AdminData element in the ODM. When connecting to the center server, authentication is performed using the center server user.

Setup Automation

Occasionally in clinical trials, there are cases where several subjects are handled at a given hospital. Requiring a system engineer to set up this system at each hospital can be costly, resulting in the operation using our system becoming more expensive than that with a regular EDC. Therefore, we developed a function to automate setup using a template master, an ODM and a study configuration file, all of which are distributed via a network from a content server to the hospitals participating in the study. The only operation required by each hospital is user management and the maintenance of the correspondence table between the universal keyword and the local code. Before starting a study, the operation staff on the hospital side need only to confirm that the ODM is correctly created using test subjects.

User Operation

Data Transcription Type

The CRF reporter is a system independent of the EMR. After logging into the system, it shows the studies for which the user has been given operation authority. When a study is selected, the subjects participating in the study are displayed. When a new subject is entered, patient identification information such as the patient ID, name, gender, date of birth, etc., are acquired from the EMR database added to a correspondence table between the subject key and the patient ID. The patient

identification information is then displayed in the subject list to help the user select the correct subject. These data, however, are not sent to the data center.

When a subject is selected, the event list for the subject is displayed. Each event includes several forms. After selecting a form, the corresponding template is displayed. By clicking on the data retrieval button in this template, the fields of the template can be populated by the data if they have already been entered by the progress note templates.

After entering all of the required items for an event, the status of the event changes to “completed”, and ODM creation becomes possible. The user checks the contents of the ODM. If no amendment is needed, they click the send button and transmit the ODM to the center server in the data center. When the center server receives the ODM correctly, the data reception status is returned to the CRF reporter.

Direct Record Type

The direct record type is currently under development and set to start operation from April 2017. The functions scheduled to be included are shown in Figure 4.

When a patient is selected in the EMR, the patient’s medical record screen is opened. On opening the progress note input screen from this point, an activation button for the CRF reporter is displayed. If the patient is a subject or a subject candidate and the user has been given operation authority, this button flashes. Clicking on this button displays the event list for the study. When a form for an event is selected, the corresponding template is displayed. The user then enters data in accordance with the items on the template. For items such as laboratory test results, the corresponding data are automatically retrieved and populated. By clicking the send button, the ODM is created and sent to the center server in the data center.

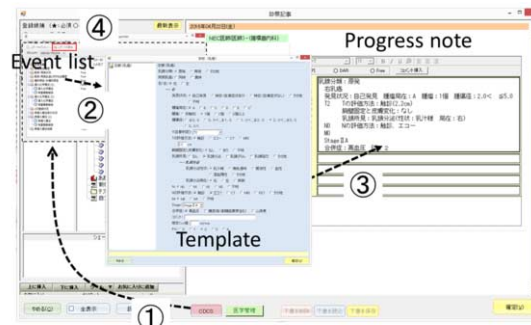


Figure 4 – Operation of Doctors when using Data Record Type

1. Open the event list of the study for the patient by clicking activation button for the CRF reporter
2. Select a form to open the corresponding template
3. Output the narrative text to record in progress note
4. Create the ODM and send it to the center server

Results

Dynamic templates have been incorporated into the EMR system developed by NEC since 2000. Thus far, more than one hundred hospitals have employed this system, and more than seventy thousand templates have been created and used in these hospitals. For example, at Osaka University Hospital, 2182 template masters have been registered, and 123,191 templates per month are used, resulting in 1,195,831 data items per month being entered.

At present, only the data transcription type CRF reporter is being used. Three hospitals have implemented this CRF

reporter in their EMR systems; however, at other hospitals, the CRF reporter is installed on a single PC for use. While the CRF reporter cannot cooperate with the EMR system in such cases, it can still be used as a terminal for usual EDCs. The ability to create CRFs offline makes this approach convenient compared with normal EDCs. Using this system, we have carried out 12 observational studies and 3 clinical trials so far. Six of these studies consisted of more than 10 events. These clinical studies have successfully collected data, and six of them have already been completed. The maximum number of the entry subjects were 1119.

Discussion

This system is a special type of EDC comprising a CRF reporter and a CDMS with a CRF receiver. The CRF reporter is built into the EMR. This model is similar to the Retrieve Form for Data Capture (RFD) model proposed by Integrating the Healthcare Enterprise (IHE) [6]. The RFD model consists of a Form Manager, Form Filler, Form Archiver, and Form Receiver. Our model consists of a content management system corresponding to a Form Manager, templates in the CRF reporter corresponding to Form Filler, a local server of the CRF reporter corresponding to a Form Archiver, and a CDMS with a CRF receiver corresponding to a Form Receiver. However, our model is not completely RFD-compliant, as we used a unique template module instead of XForms for the form filler and REST communication instead of SOAP for communication between Form Filler and Form Receiver. However, several challenges to the EDC and EHR interoperable systems have been reported, and a completed system based on the RFD model has not yet been developed.

Regarding the integration of EHRs and EDCs, one report described a system that retrieves the data elements of patient care-centric template in use in the EHR to pre-populate the CRFs, but only 13.4% of the data elements in the CRFs were able to be retrieved [7]. In our model, even with the data transcription type of CRF reporter, a template used for patient care is purposefully created to allow users to enter data for a given study; as such, almost all data collected in the study can be retrieved. In addition, since narrative expression-style text generated from the data are entered as a record of the progress note and doctors can freely add supplementary text to the progress note, the note can be treated as a medical record as well.

For registries, various patient data such as the symptoms, findings, laboratory test results, image examination reports, and treatments, are entered. In the data transcription type, a template for the progress note can be created freely at each hospital. When registering a single case in multiple registries, if a template with unified items in the case report forms for multiple registries would be created, with the initially entered data then retrieved by the templates for the respective CRFs and the completed CRFs sent to the corresponding data centers.

For clinical and prospective observational studies, certain forms must be input for each visit, and some forms are used across several visits. If such studies use data transcription type CRF reporters, the template corresponding to each visit is not shown and data input cannot be guided. Another problem is that retrieving the data for the same form from different visits using the corresponding CRF reporter form is complicated. In contrast, when using the direct record type of CRF reporter, the events and forms to be entered are displayed and their statuses shown, guiding doctors to input data into the proper template for each visit. The preparation of the CRF reporter for the direct record type is easier than that for the data transcription type. As

such, for prospective clinical studies, the direct record type is more suitable.

This system saves labor through the simultaneous entry of the medical record and the CRF as well as the automatic retrieval of the data recorded in the EMR. The former is an essential function of this system, but the latter depends on the interface function between the template and EMR. The more easily various data can be retrieved from the EMR database, the more efficient clinical research becomes. Details of this function of our system were reported in MIE in 2016 [8].

When integrating EDCs and EHRs, locus of responsibility should be clear [9]. With this system configuration, the CRF reporter is a investigator side system. Therefore, we believe that the validation for clinical research is required within the scope of using a CRF reporter. For data transcription type CRF reporters, since the system automatically retrieves the data and applies it to the template of the CRF reporter, the mechanism of correct retrieval must be validated. For the direct record type as well, we must also validate the automatic retrieval function for laboratory test data and similar values; however, the data retrieval range is narrow compared to the data transcription type. For the direct data record type, since data is output to the medical chart, we must validate whether or not this output is correctly recorded in the EMR. From the viewpoint of the locus of responsibility, the investigator side should be responsible for the CRF reporter and for directly transferring data between the CRF reporter and EMR, regardless of data transcription type or direct record type.

For the data transcription type, the doctor may describe the record in free text in the progress note, and the CRC manually transcribes the data to the template of the CRF reporter. Even when doctors input data using the template in the progress note, after the CRC transcribes the data into the CRF, the doctor may change the data using the template, which may result in inconsistencies between the CRF and EMR record. Therefore, interaction with the data transcription type CRF reporter requires an SDV, as in ordinary EDCs. In contrast, in the direct record type, the contents entered by the CRF reporter are forcibly added to the medical record, so it matches the contents of the medical record. Also, when revising, since the template of the CRF reporter is opened and modified, inconsistency does not occur. Therefore, an SDV is unnecessary when the direct record type CRF reporter is used. This further underscores the suitability of the direct record type over the data transcription type for clinical trials.

The integration of EDCs and EHRs relies heavily on the unification of terms. However, using the direct record type of reporter essentially creates a system closed to the EDC, so there is no need to match the term code of ItemData with the EMR side. In contrast, when using the data transcription type of reporter, the item code of the template in the progress note must be designated in the field of the template of the CRF reporter in order to retrieve data. We assume that a template for the progress note is additionally created for clinical research, and doctors purposely select that template if a patient is participating in a given study. It is therefore not absolutely necessary to unify item codes of both templates. However, if the CRF reporter would be set independently from the creation of the templates in the progress note, a universal keyword must be wrote as a concept code of a given element of the template in the progress note.

It has often been said that a unified vocabulary for medical care and clinical research is prerequisite for the interoperability of EHRs and EDCs [10;11]. However, compiling a complete dictionary of terms in both medical care and clinical research is an endless task. Thus, if the creation of this dictionary was premised, the integration of EHRs and EDCs would be

realizable. In our system, while establishing a list of unified keywords would be useful, it is not necessary. We feel that this element makes our system practical for clinical research support.

Because the CRF reporter is an independent component of the EMR, it can be incorporated in any EMR system. We have so far incorporated it in the EMR developed by NEC and another by IBM. With accordance of the communication procedure between CRF reporters and the receiver, a different type of CRF reporter comes into effect. Fujitsu is trying to develop another type of CRF reporter which can communicate with the same center server. By incorporating the CRF reporters in the EMRs in the hospitals, we are creating a hospital cluster for networked clinical research, which will consist of more than 16 hospitals in Osaka in the following years.

Conclusion

We developed a CRF reporter to integrate EHRs and EDCs. The CRF reporter incorporated in an EMR system can retrieve data recorded in the EMR, generate an ODM, and transmit the ODM to the CDMS. With the data transcription type of CRF reporter, doctors enter data using a template for the progress note, and the data are then transmitted to the template for the CRF reporter. In the direct record type of reporter, templates of the CRF reporter open from the progress note, generating narrative text to make a record in the progress note. This system can save both labor and financial costs in clinical research.

Acknowledgements

This research is supported by the Project of ICT infrastructure establishment for clinical and medical research from Japan Agency for Medical Research and Development, AMED.

References

- [1] C. Ohmann and W. Kuchinke, Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration, *Methods Inf Med* **48**(1) (2009), 45-54.
- [2] CDISC Electronic Source Data Interchange (eSDI) Group, Leveraging the CDISC standards to facilitate the use of electronic source data within clinical trials version, November 20, 2008. < <http://www.cdisc.org/stuff/contentmgr/files/0/2f6eca8f0df7caac5bbd4fadfd76d575/miscdocs/esdi.pdf> >
- [3] Y. Matsumura, A. Hattori, S. Manabe, T. Takeda, D. Takahashi, Y. Yamamoto, T. Murata, and N. Mihara, Interconnection of Electronic Medical Record with Clinical Data Management System by CDISC ODM, *Stud Health Technol Inform* **205** (2014), 868-872.
- [4] Clinical Data Interchange Standards Consortium (CDISC) , 15907 Two Rivers Cove, Austin, Texas, USA; ©2008. Standard available from: <http://www.cdisc.org/models/odm/v1.3/index.html> (cited Apr 20, 2009).
- [5] Y. Matsumura, S. Kuwata, Y. Yamamoto, K. Izumi, Y. Okada, M. Hazumi, S. Yoshimoto, T. Mineno, M. Nagahama, A. Fujii, and H. Takeda, Template-based data entry for general description in medical records and data transfer to data warehouse for analysis, *Stud Health Technol Inform* **129** (Pt 1) (2007), 412-416.
- [6] ITI Technical Committee. IHE IT Infrastructure Technical Framework Supplement Retrieve Form for Data Capture (RFD) Trial Implementation. August 19 (2011), <http://www.ihe.net/technical_framework/upload/ihe_iti_suppl_rfd_rev2-2_ti_2011-08-19.pdf>
- [7] A. El Fadly, B. Rance, N. Lucas, C. Mead, G. Chatellier, P.Y. Lastic, M.C. Jaulent, and C. Daniel, Integrating clinical research with the Healthcare Enterprise: from the RE-USE project to the EHR4CR platform, *J Biomed Inform Dec*; **44** Suppl 1(2011), S94-102.
- [8] Y. Matsumura, A. Hattori, S. Manabe, T. Tsuda, T. Takeda, K. Okada, T. Murata, and N. Mihara, A strategy for reusing the data of electronic medical record systems for clinical research, *Stud Health Technol Inform* **228** (2016), 297-301.
- [9] J.D. Mestler, C. Celingant, S. Bishop, R. Perkins, and M. Rocca, EHR/CR Functional Profile Working Group. EHR/CR Functional Profile, Electronic Health Records/Clinical Research. Committee Level Ballot Release 1, March (2008), http://www.eurorec.org/files/_filesPublic/ehworkshop/2008/HL%20EHR%20Functional%20Profile%20Rel%201.pdf

- [10] W. Xu, Z. Guan, J. Sun, Z. Wang, and Y. Geng, Development of an open metadata schema for prospective clinical research (openPCR) in China, *Methods Inf Med* **53**(1) (2014), 39-46.
- [11] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, and J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare, *Database (Oxford)* 2016 Feb 11, 1-9.

Address for correspondence

Yasushi Matsumura MD, PhD
2-15, Yamada-Oka, Suita, Osaka 565-0871, JAPAN
Phone: +81-6-6879-5900, Fax: +81-6-6879-5903
e-mail: matumura@hp-info.med.osaka-u.ac.jp

Scholarly Information Extraction Is Going to Make a Quantum Leap with PubMed Central (PMC)[®] —But Moving from Abstracts to Full Texts Seems Harder than Expected

Franz Matthies and Udo Hahn

Jena University Language & Information Engineering (JULIE) Lab
Friedrich-Schiller-Universität Jena, Jena 07743, Germany

Abstract

With the increasing availability of complete full texts (journal articles), rather than their surrogates (titles, abstracts), as resources for text analytics, entirely new opportunities arise for information extraction and text mining from scholarly publications. Yet, we gathered evidence that a range of problems are encountered for full-text processing when biomedical text analytics simply reuse existing NLP pipelines which were developed on the basis of abstracts (rather than full texts). We conducted experiments with four different relation extraction engines all of which were top performers in previous BioNLP Event Extraction Challenges. We found that abstract-trained engines loose up to 6.6% F-score points when run on full-text data. Hence, the reuse of existing abstract-based NLP software in a full-text scenario is considered harmful because of heavy performance losses. Given the current lack of annotated full-text resources to train on, our study quantifies the price paid for this short cut.

Keywords:

Natural Language Processing; Information Storage and Retrieval; Information Extraction

Introduction

While abstracts contain only the major results of the corresponding journal article in a highly condensed manner, the full-text body of scholarly publications makes accessible all single pieces of information from scientific studies. Since the demand for this maximum level of information is also high in the life sciences, the NLM launched a full-text collection for a subset of PUBMED abstracts, PUBMED CENTRAL (PMC),¹ which currently (April 2017) archives 4.3 million articles and is growing at high speed. Given such a bulky resource, unlimited access to the information contained in scholarly full texts seems both a realistic and rewarding goal, since a great deal of relevant biomedical information is only contained in the full-text portions of scientific articles and is not mentioned at all in the corresponding abstracts of the full texts [1–3] – so the information gain from processing full texts can be enormous.

Tanabe and Wilbur [4] were the first to hint at technical problems of dealing with special non-ASCII characters, tables and figures embedded in full texts. Going beyond low-level technicalities, within the BIOCREATIVE Gene Normalization Challenge, for the first time, evidence was gathered that performance dropped significantly when tested on full texts [5] instead of abstracts [6].

Cohen *et al.* [7] substantiated these warnings that the processing of full texts will be more than challenging by investigating their intrinsic properties. They conducted an empirical study in which they investigated different structural and linguistic properties of abstracts and their corresponding full texts. They found longer sentences in the full texts than in the abstracts and much heavier use of parenthesized material (e.g., abbreviations, citations, data values, figure/table pointers, etc.) in the full texts than in the abstracts. Both phenomena make full texts much harder to parse than abstracts. Syntax-wise they also gathered evidence that the incidence of conjunctions, passives, pronominal anaphora, as well as sentence complexity/readability were markedly different from full texts to abstracts. However, syntactic parsing using the Stanford Lexicalized Parser yielded no statistically relevant difference between both genres (taking ParsEval's metrics for bracket recall and tag accuracy). Yet, POS tagging was more accurate in abstracts than in full texts. Analysis errors caused by the increased syntactic complexity of full text articles (compared with abstracts) were also recognized by Tudor *et al.* [8]. McIntosh and Curran [9] further point out that coreference relations (anaphora) between sentences play an important role in full texts. Semantics-wise, Cohen *et al.* found that the distribution of named entities such as genes/proteins, mutations, drugs, etc. differed between the two text genres as well.

Most important for our own work, Cohen *et al.* also investigated the impact these differences had on the performance of information extraction tools. They found that three common gene mention recognition systems (BANNER, ABNER and LINGPIPE) performed much worse on full texts than they did on abstracts – F-measures were generally about 10 points higher on the abstracts than on the full text portions. Consequently, the authors advocated retraining models on full texts.

If this suggestion is taken seriously, annotated full-text corpora have to be supplied. Given the large volume of utterances contained in full texts, this is an enormously resource-consumptive task. Cohen *et al.*'s study is based on their self-developed CRAFT full text corpus [10] which reflects the findings for full text phenomena from [7]. It consists of 97 (67 publicly released) full texts, plus corresponding abstracts (comprising nearly 31k sentences or 800k text tokens), dealing with mouse genomics. This corpus was annotated for part-of-speech (POS) and syntactic parsing data, as well as proteins/genes as named entities [11].

¹ <https://www.ncbi.nlm.nih.gov/pmc/>

Table 1 - Overview of relation extraction systems used for our experiments and their ranking in the BioNLP Shared Task 2013 on the GENIA Task (in parentheses, results from the BioNLP Shared Task 2009, if applicable)

Name	Type	Ranking		Group	Reference
TEES	ML (SVM)	2.	(1.)	Turku, FI	[24], [25]
BIOSEM	ML + rule-based	3.	(n.a.)	Amsterdam/Rotterdam, NL	[26], [27]
HDS4NLP	ML (SVM)	6./1.	(n.a.)	Compiègne, FR	[28]
JREX	ML (SVM, MAXENT)	n.a.	(2.)	Jena, DE	[14], [29]

The authors carefully discuss the large variation in testing conditions, e.g., the types of named entity taggers (BANNER, ABNER and LINGPIPE), the types of models being considered (trained on the abstract-based BIOCREATIVE II, NLPBA and GENIA and the full-text-based CRAFT corpora where, in addition, the full set is distinguished from a development set), different definitions of the named entity type ‘protein’ and different matching criteria for gene mentions). Still, the CRAFT study reveals inconsistent results for the ‘gene’ named entity recognition problem. For the LINGPIPE system, gains for retraining on the full-text corpus amount to 0.18 F-score points, but ABNER’s performance drops after retraining (increased precision comes at the cost of substantial recall losses). BANNER’s results are already competitive with the out-of-the-box model so that no retraining on CRAFT-style full text corpora is required. The authors summarize their results that “retraining the gene mention recognition systems unfortunately did not show much improvement” [10, p. 21] but anticipate a significant improvement if the learning problem will be rephrased (see also [10, p. 21]). Note that CRAFT is not annotated for any relational information so that no empirical data could be collected for this task from this resource. The CRAFT study also measured the impact of sentence splitting, tokenization, POS tagging and syntactic parsing relative to the two text genres—abstracts and full texts. For parsing, e.g., substantial improvements of CRAFT-trained models over standard (non-biomedical) English models were found.

Another full-text corpus, ID, remedies the lack of biomedical relation annotations. It was supplied for the series of BioNLP Event Extraction Challenges starting in 2009, with the second round in 2011 [12]. ID contains 30 full texts (more than 5k sentences, 150k text tokens) annotated for biomolecular mechanisms of infectious diseases that involve associations between multiple types of molecular entities, disease-causing microorganisms and other organisms undergoing the diseases. With more than 13k named entities and 4,15k events the entity count is comparable to the 2009 Challenge, whereas the event count is only approximately one third of the Shared Task 2009 data. On this data, the top-performing systems scored at almost the same level as in 2009 (where abstracts were analyzed), with the winner system (FAUST) peaking on 56% F-score (trained and evaluated on sections of ID full texts).

Since the biomedical NLP community has developed a battery of well-engineered analytic engines—starting from domain-adapted tokenizers, POS taggers and parsers to domain-specific named entity taggers and relation extractors—one is tempted to simply reuse these tools and composite pipelines on full texts in order to unlock the vast amount of still hidden information. The fact that almost all of them were developed and fine-tuned on abstracts as textual training data does not seem to be an issue here. Instead, we claim that moving from the abstract to the full-text level of analysis is by no means a free lunch. We rather stipulate that classifiers trained on abstracts (irrespective of whether they deal with named entities or relations between them) will drop significantly in performance when run on corresponding full texts due to the increased level of linguistic, structural and conceptual complexity in the latter. Thus, the potential cheap benefit of making full texts available, instead of (informationally much

poorer) abstracts, is likely to disappear. —But at what rate? And, is this rate, when quantified, tolerable or not?

In this paper, we deal with the problem of trading off abstract vs. full-text processing for the life sciences domain, with focus on relation extraction. We thus study the effects of text genres on system performance by switching between abstract and full-text documents. We substantiate our claims by running four prominent relation extraction systems, which were top-performers in the most recent BioNLP 2013 Event Extraction Challenge [13]. This selection of systems should guarantee that despite the small ‘sample size’ of the four classifiers the results we achieve might cautiously be generalized for a much larger class of relation taggers in the biomedical domain.

Methods

In this section, we describe the experimental set-up of our work. First, we introduce the systems we used for relation extraction. Then, we describe the full-text resources for our experiments. From a methodological perspective, a general observation can be made. The best-performing BIOINLP systems (cf. the top performers in challenge competitions such as BIOCREATIVE [15-17], BIOINLP Shared Task [18,19], i2B2 [20,21], or DDI [22,23]) either exclusively rely on some (semi-)supervised form of machine learning (ML) techniques, or combine ML with rule- or dictionary-based approaches in terms of hybrid systems. The ML systems (or ML portions of hybrid systems) being used are thus highly dependent on their textual input for training, i.e., subsets of PUBMED (MEDLINE) abstracts annotated by humans, since they constitute the gold standards both for training and evaluating these classifiers. Unsupervised ML systems are rare and, if running against competitor systems, are usually outperformed by supervised approaches in the challenges.

Relation Extractors

For our experiments, we used four systems that performed exceptionally well in previous BioNLP Event Extraction Challenges. Another criterion of choice was technical in nature—the code of the systems should be easily accessible and processable without much effort in our computing environment. The latter is important because technical portability of systems is an indicator for the extramural reproducibility of results. In the following, we briefly summarize each system and point out main differences among them and their approaches to relation extraction. For a more in-depth description, we kindly refer to the respective papers cited in Table 1.

TEES

The TEES system [24] was successful in all three BioNLP Shared Tasks, achieving the first rank in 2009 and the first rank in half of the subtasks in 2011 and 2013, respectively. Since the team has also provided the code as an open source project² to the scientific community, it was clearly the first choice for our experiments.

TEES approaches the task of relation extraction in a linear fashion, by first detecting potential triggers in a text and, in

² <https://github.com/jbjorne/TEES>

the next step, defining valid edge candidates between entities: triggers or arguments. At this point, overlapping events are possible—that is, they share the same trigger node. This ‘merged event graph’ is unmerged in the next step. A Support Vector Machine (SVM) sequentially and independently solves all these steps with a linear kernel handling them as multi-class classification problems.

BIOSEM

The BIOSEM system³ [26] stands out from the other three systems we consider here because of two interesting facts. Firstly, in the training phase, it learns rules for relation representations by means of a semantic and syntactic feature list and, secondly, it refrains from deep syntactic parsing, but rather builds its structured representation of the text from the output of a shallow parse (i.e., chunks) only. Among other decisions, this leads to an outstanding computational efficiency, paving the way to relation extraction on a larger scale. The authors estimate their system to be around 170–230 times faster for completely extracting an event from a sentence compared to state-of-the-art ML-based systems. (see [26] “Computational Performance”). The usage of rules also yields a superior precision (between 60–70% depending on the set-up) compared to the other systems we consider here.

HDS4NLP

In the official results for the GENIA task, HDS4NLP [28] only achieved rank 6 (with 43.03 F-score). However, its developers identified a serious bug after disclosure of the test results, and, after fixing it, their system reportedly achieved an F-score of 51.15, outperforming the top ranked competitor EVEX (50.97 F-score).⁴ Since they trained the model for producing the results for the 2013 Shared Task on all documents from both the development and the training sets of the BioNLP 2011 and 2013 GE tasks, the numbers we present here differ from the ones they reported, as we only use 2013 GE task documents. However, we still achieve a much higher F-score (47.81) than the official figures, putting them among the top five performers of the challenge. Also, HDS4NLP outperforms the other three systems when evaluated on terms of the BioNLP 2009 Shared Task with 54.37 F-score by at least 2.3 points.

In comparison to the other systems used for our experiments, HDS4NLP tackles relation extraction by training a model that directly extracts pairwise structured events (and the event type they belong to) of the form (*trigger*, *argument*) rather than relying on a sequential technique (i.e., extracting triggers first and then looking for applicable arguments). A point the authors make why this might be beneficial is that the usual approach of detecting triggers in isolation could lead to contextual information loss. Running an SVM (as implemented through PYTHON’s SCIKIT-LEARN environment)⁵ on the sentence level addresses the problem of classifying candidate-argument pairs [26]—taken from the cross product of the sets of possible trigger tokens and arguments—in a one-vs.-all set-up. We will elaborate on this point in the Results Section.

JREX

The JREX system, developed in our lab, participated only in the BioNLP 2009 GE task [14] where it ranked on 2nd place among 24 teams. Since then it has undergone almost no updates. This stagnation is clearly reflected in the performance figures compared to the other three systems used in our

experiments which are/were under active development. JREX incorporates manually curated dictionaries and ML methodologies (SVM, MAXENT) to sort out associated event triggers and arguments on *trimmed* dependency graph structures, the latter being simplified dependency structures from which representational ‘noise’ has been eliminated (cf. [14, 29]).

Text Corpora

The textual resources we exploited were taken from the text repository of the 2013 Event Extraction Challenge for the full texts and the 2011 repository for the abstracts (cf. Table 2 for a quantitative breakdown). For the former to get a reasonable amount of material, we used training and development set, whereas for the latter we only used the train set for training as to keep the size approximately on a par with that of the full text collection.

Table 2 - Overview of Text Corpora; mix = train \cup dev

Items	Abstract		Full Text		
	train	test	train	mix	test
Documents	800	260	10	20	14
Sentences	7,449	2,447	2,438	5,165	3,204
Words	176,146	57,367	54,938	112,845	75,144
Events	8,597	3,182	2,817	6,016	3,270

This set-up results in performance figures for the systems that differ from those reported in the 2009 Shared Task, since they used both development and training sets as input for training. Differences in the results when replicating the Shared Task 2013 set-up (compared to the original scores) can be accredited to the fact that all the systems used 2011 (abstracts, as well) and 2013 resources for training.

Events

In order to give an idea of the quantitative scope of the text collections we used, in Table 3, we distinguish the Abstract and Full Text corpus, with counts of all

- *simple* unary relations which refer to all events constituted only of an *event trigger* (a sequence of tokens indicating an event mention) and an *argument*, a protein or gene,
- the binary *Binding* relation which can have two *arguments* (both a gene or a protein),
- *Regulation* relations, which – besides proteins or genes – can also have other events as *arguments*.

Results

In this section, we present the results the four systems achieved for all possible combinations of training and evaluation data, i.e., the cross-product of the abstract (AB) and full text (FT) material. Hence, we:

- trained models on the Abstract corpus and evaluated them on Abstracts (**AB on AB**),

Table 3 - Relation Type Counts (Test set) – The numbers for Full Texts do not incorporate relation counts for Protein modification and Ubiquitination.

Relation Type	Abstract	Full Text
Simple Relations	1,182	993
Binding Relations	347	333
Σ Relations	1,529	1,326
Regulation Relations	1,653	1,944
Total	3,182	3,270

³ In consultation with the main developer their source code was adapted to be fully compatible with the UIMA framework and is available from:

<https://github.com/JULIELab/jcore-dependencies/tree/master/biosem-event-extractor>

⁴ Available from <https://github.com/XiaoLiuAI/RUPEE>

⁵ <http://scikit-learn.org>

Table 4 - Results for all conditions in Recall/Precision/F-Score triples; bold values mark the best value for the respective set-up

Name	AB on AB	AB on FT	FT on FT	FT on AB
TEES	(46.67/56.74/51.40)	(39.47/50.64/45.01)	(43.17/57.30/49.24)	(40.19/51.57/45.18)
BIOSEM	(41.39/70.17/52.07)	(37.11/63.05/46.72)	(38.41/67.65/49.00)	(37.12/71.36/48.83)
HDS4NLP	(46.32/65.80/54.37)	(43.77/59.12/50.30)	(42.84/54.09/47.81)	(39.53/54.36/45.78)
JREX	(39.75/51.97/45.05)	(37.20/44.80/40.65)	(37.69/48.03/42.23)	(37.15/52.51/43.51)

Table 5 - Deltas in Recall/Precision/F-Score triples for comparison between different columns of Table 4

Name	AB on AB			FT on FT	
	FT on FT	AB on FT	FT on AB	FT on AB	AB on FT
TEES	(-3.50/+0.56/-2.16)	(-7.20/-6.10/-6.39)	(-6.48/-5.17/-6.22)	(-2.98/-5.72/-4.06)	(-3.70/-6.66/-4.23)
BIOSEM	(-2.98/-2.52/-3.07)	(-4.28/-7.12/-5.35)	(-4.27/+1.19/-3.24)	(-1.29/+3.71/-0.17)	(-1.30/-4.60/-2.28)
HDS4NLP	(-3.48/-11.71/-6.56)	(-2.55/-6.68/-4.07)	(-6.79/-11.44/-8.59)	(-3.31/+0.27/-2.03)	(+0.93/+5.03/-2.49)
JREX	(-2.06/-3.94/-2.82)	(-2.55/-7.17/-4.40)	(-2.60/+0.54/-1.54)	(-0.54/+4.48/+1.28)	(-0.49/-3.23/-1.58)

- trained models on the Abstract corpus and evaluated them on Full Texts (**AB on FT**),
- trained models on the Full Text corpus and evaluated them on Abstracts (**FT on AB**),
- trained models on the Full Text corpus and evaluated them on Full Texts (**FT on FT**).

Table 4 shows the resulting figures in terms of Recall/Precision/F-Score triples, whereas Table 5 depicts the deltas (differences) for different condition comparisons across set-ups. When simply comparing the systems' intra-collection performance (cf. columns **AB on AB** vs. **FT on FT**), we already see an F-Score difference (Δ) in favor of the abstracts, where TEES' FT performance is the most robust (only 2.16 F-score points loss compared with AB), whereas HDS4NLP turns out to be the most fragile system (6.56 F-score points loss compared with AB).

This is an early indicator that relations are generally harder to extract from full texts than from abstracts. However, due to the difference in the relation count between the training material for Abstracts and Full Texts this conclusion is still preliminary and will be treated with some reservation. We will come back to this issue in the Discussion Section.

A noteworthy performance drop, ranging from approximately 4.1 to 6.4 points in F-Score, can be observed consistently for all systems when evaluating abstract-trained models on full text compared to a scenario when these models were tested on abstracts only (cf. columns **AB on AB** vs. **AB on FT**). These deltas are already strong despite the fact that we could not test for statistical significance (see Section Discussion).

In combination, both result sets already indicate that *all* systems fail to capture relation structure encodings that are present in the full text but not in the abstract. This view might be further encouraged by looking at the performance of Full Text models evaluated on Abstract and Full Text, respectively (cf. columns **FT on FT** vs. **FT on AB**). Not only do these results show no such clear picture (the deltas are much smaller), but also does the JREX system goes the other way round (F: +1.3).

Discussion

There are some caveats that need to be made explicitly for the experimental set-up we have defined:

- Text Genre Mix-up.** The collection we here referred to as "Full Text" does include the abstracts. We could not change this mix-up because we, obviously, do not have any access to the test set. We could have done this separation for the training phase, but this would have led to the elimination of roughly about 400 relations.
- Training Set Imbalance.** The textual material used for training the Full Text models is lower in size than that for training the Abstract models (approximately 6,000 vs.

8,600 events, respectively). This could be seen as a problem for comparability of the results. Yet, we still have the strong figures where Abstract models were tested on both AB and FT, and we did further experiments on both JREX and BIOSEM where these systems were subsequently tested with more and more training material for both AB and FT on both AB and FT (data not shown). In order to get an approximate increase of 2.5 F-Score for AB training material (regardless of the testing material; AB or FT) one needs a 100 % increase of material (from 4.300 to 8.600 events; the full train set size)

- Statistical Significance.** We refrained from testing the statistical significance of our results. The major issue we faced is the data sparseness of full texts (another strong point for more annotated full texts): to perform said test under minimal requirements, we would need to split the full text material in such a way that either the training or test set would hold too few events to be representative.
- Error Analysis.** To provide a thorough error analysis, we would need to deal with 16 cases (all four systems with all four combinations of AB and FT for training and evaluation). We leave this discussion for a companion paper.

The outlier results for the HDS4NLP scores also deserve several remarks: The system

- achieves the highest F-Scores when trained and evaluated on abstracts—54.37; see Table 4 (**AB on AB**),
- yet has the lowest difference when these AB models are evaluated on full texts instead ($\Delta = 4.07$); see Table 5, column 2,
- drops the highest ($\Delta = -8.59$) when FT-trained instead of AB-trained models are used for evaluation on the AB test sets; see Table 5, column 3
- and further drops the highest when comparing **AB-AB** vs. **FT-FT** ($\Delta = -6.56$); see Table 5, column 1.

We tend to explain this special role of the HDS4NLP system as a result of taking an entirely different methodological approach than the other three systems—HDS4NLP directly extracts pairwise structured events (and the event type they belong to) of the form (*trigger*, *argument*) instead of using a linear technique (i.e., extracting triggers first and then looking for applicable arguments). From the three competitive systems, HDS4NLP performs the worst in a solely **FT on FT** set-up so that these preliminary analyses could lead to the cautious conclusion that the HDS4NLP system is best (and better) suited to utilize abstracts as training material.

The exceptions coming from the JREX system (see its inconsistent increase in performance when Full Text models are evaluated on Abstract data, Table 5, column 4) are completely overshadowed by a substantial performance penalty in comparison with all other systems. This is clearly an effect of lacking maintenance over the past five years.

Conclusions

We focused in this paper on the implications of running established tooling and pipelines, developed on scientific abstracts as training data, on scholarly full text data from the life sciences. Our experiments reveal that F-score losses up until 6.6 F-Score points have to be anticipated. Abstracting away from the specific system particularities, there is a consistent trend for performance degradation when abstract-trained models are transferred without changes to full texts.

The apparent solution—providing annotated full-text corpora—is costly and resource-intensive. Currently, only one corpus (ID) provides relation encodings for full texts, whereas another one contains only highly specific named entity encodings (CRAFT). Given the resource-density in the life sciences field, distant supervision (via database contents) might be a reasonable alternative [30], while moving to unsupervised relation extraction is likely to be accompanied with an additional penalty in terms of sliding F-scores. Hence, approaches decomposing full texts into more digestible text portions (section and paragraph segmentation at the macro-level of text analysis; sentence simplification at the micro level) might be worthwhile to be taken into account.

References

- [1] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky, Genies: A natural language processing system for the extraction of molecular pathways from journal articles, *Bioinformatics* **17** (Suppl 1) (2001), S74–S82.
- [2] P. Shah, C. Perez-Iratxeta, P. Bork, and M. Andrade, Information extraction from full text scientific articles: Where are the keywords? *BMC Bioinformatics* **4**(20) (2003).
- [3] D. Corney, B. Buxton, W. Langdon, and D. Jones, BioRAT: Extracting biological information from full-length papers, *Bioinformatics* **20**(17) (2004), 3206–3213.
- [4] L. K. Tanabe and W. J. Wilbur, Tagging gene and protein names in full text articles, in *BioNLP 2002 — Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain @ ACL 2002* (2002), 9–13.
- [5] Z. Lu *et al.*, The Gene Normalization Task in BioCreative III, *BMC Bioinformatics* **12**(S8) (2011), S2.
- [6] A. A. Morgan *et al.*, Overview of BioCreative II Gene Normalization, *Genome Biol* **9**(S2) (2008), S3.
- [7] K. B. Cohen, H. L. Johnson, K. M. Verspoor, C. Roeder, and L. E. Hunter, The structural and content aspects of abstracts versus bodies of full text journal articles are different, *BMC Bioinformatics* **11** (2010), 492.
- [8] C. O. Tudor, K. E. Ross, G. Li, K. Vijay-Shanker, C. H. Wu, and C. N. Arighi, Construction of phosphorylation interaction networks by text mining of full-length articles using the eFIP system, *Database – J. Biol. Databases Curation* #bav020, 2015.
- [9] T. McIntosh and J. R. Curran, Challenges for automatically extracting molecular interactions from full-text articles, *BMC Bioinformatics* **10** (2009), 311.
- [10] K. M. Verspoor *et al.*, A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools, *BMC Bioinformatics* **13** (2012), 207.
- [11] M. Bada *et al.*, Concept annotation in the CRAFT corpus, *BMC Bioinformatics* **13** (2012), 161.
- [12] S. Pyysalo *et al.*, Overview of the ID, EPI and REL tasks of BioNLP Shared Task 2011, *BMC Bioinformatics* **13** (S11) (2012), S2.
- [13] J.-D. Kim *et al.*, The Genia Event and Protein Coreference tasks of the BioNLP Shared Task 2011, *BMC Bioinformatics* **13** (Suppl 11) (2012), S1.
- [14] E. Buyko, E. Faessler, J. Wermter, and U. Hahn, Event extraction from trimmed dependency graphs, in *BioNLP – Proceedings of the Workshop on Biomedical Language Processing @ NAACL-HLT 2009. Shared Task on Event Extraction* (2009) 19–27.
- [15] C.-H. Wei *et al.*, Assessing the state of the art in biomedical relation extraction: Overview of the BioCreative V chemical-disease relation (CDR) task, *Database – J Biol Databases Curation* baw032, 2016.
- [16] C. N. Arighi *et al.*, BioCreative-IV virtual issue, *Database – J Biol Databases Curation*, bau039, 2014.
- [17] C. H. Wu *et al.*, BioCreative-2012 virtual issue, *Database – J Biol Databases Curation*, bas049, 2012.
- [18] C. Nédellec *et al.*, Overview of BioNLP Shared Task 2013, in *Proceedings of the BioNLP Shared Task 2013 Workshop* (2013), 1–7.
- [19] J.-D. Kim, S. Pyysalo, T. Ohta, R. Bossy, N. Nguyen, and J.-i. Tsujii, Overview of BioNLP Shared Task 2011, in *Proceedings of BioNLP Shared Task 2011 Workshop*, (2011), 1–6.
- [20] Ö. Uzuner and A. Stubbs, Practical applications for natural language processing in clinical research: The 2014 i2b2/UTHealth Shared Tasks, *J Biomed Inform* **58** (S1–S5) (2015).
- [21] Ö. Uzuner, B. R. South, S. Shen, and S. L. DuVall, 2010 i2b2/VA Challenge on concepts, assertions, and relations in clinical text, *JAMIA* **18**(5) (2011), 552–556.
- [22] I. Segura-Bedmar, P. Martínez, and M. Herrero-Zazo, Lessons learnt from the DDIEExtraction-2013 Shared Task, *J Biomed Inform* **51** (2014), 152–164.
- [23] I. Segura-Bedmar, P. Martínez, and D. Sánchez-Cisneros, The 1st DDIEExtraction-2011 Challenge Task: Extraction of drug-drug interactions from biomedical texts, in *DDIEExtraction – Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction 2011. SEPLN 2011 Satellite Workshop* (2011), 1–9.
- [24] J. Björne and T. Salakoski, Tees 2.1: Automated annotation scheme learning in the BioNLP 2013 Shared Task, in *Proceedings of the BioNLP Shared Task 2013 Workshop* (2013), 16–25.
- [25] J. Björne, F. Ginter, and T. Salakoski, University of Turku in the BioNLP’11 Shared Task, *BMC Bioinformatics* **13**(11) (2012), 1–13.
- [26] Q.-C. Bui, D. Campos, E. M. van Mulligen, and J. A. Kors, A fast rule-based approach for biomedical event extraction, in *Proceedings of the BioNLP Shared Task 2013 Workshop* (2013), 104–108.
- [27] Q.-C. Bui and P. M. A. Sloot, A robust approach to extract biomedical events from literature, *Bioinformatics* **28**(20) (2012), 2654–2661.
- [28] X. Liu, A. Bordes, and Y. Grandvalet, Biomedical event extraction by multi-class classification of pairs of text entities, in *Proceedings of the BioNLP Shared Task 2013 Workshop* (2013), 45–49.
- [29] E. Buyko, E. Faessler, J. Wermter, and U. Hahn, Syntactic simplification and semantic enrichment: Trimming dependency graphs for event extraction, *Computational Intelligence* **27**(4) (2011), 610–644.
- [30] E. K. Mallory, C. Zhang, C. Ré, and R. B. Altman, Large-scale extraction of gene interactions from full-text literature using DeepDive, *Bioinformatics* **32**(1) (2016), 106–113.

Address for correspondence

Franz Matthies: franz.matthies@uni-jena.de

Can Sonification Become a Useful Tool for Medical Data Representation?

George I Mihalas^{a,c,d}, Minodora Andor^b, Anca Tudor^a, Sorin Paralescu^b

^a Department of Medical Informatics, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

^b Department of Medical Semiotics, Victor Babes University of Medicine and Pharmacy, Timisoara, Romania

^c Commission of Medical Informatics and Data Protection, Romanian Academy of Medical Sciences, Bucharest, Romania

^d Department of Biophysics, Vasile Goldis West University, Arad, Romania

Abstract

This 'vision' paper refers to sonification – a novel method to represent data by sounds. A short theoretical background comprises the main features to attach sound to a set of data – how to map the correspondence between the sound parameters (pitch, duration) and the initial set of data. The classification of sonification methods is followed by a description of sound display tools – tempolenses and artifacts (saccadic display or loudness variations). The Results section comprises examples of sonification performed by our team: heart rate (HR), ECG signals, HR variations during exercise, including warning procedures. The procedure to evaluate the discriminant power of various sonification algorithms is then described. As a 'vision' paper, the most important part is not represented by the results, but the potential future developments, presented in the Discussion section, which starts with a critical view of the present state and presents future potential applications of sonification in medicine.

Keywords:

Sound; Heart Rate; Electrocardiography

Introduction

There are several intuitive means for data representation in science, developed for a better understanding of various phenomena, mostly based on performances of visual perception. Indeed, all visualization procedures have proved their capacity to give us a deeper insight of the processes and even reveal interesting features. But, the auditory system is also feeding our brains with different kinds of information, a way still insufficiently explored. Yet, some attempts have been made to develop such procedures, yielding a new method for data representation – sonification. It is defined as "the use of non-speech audio to transform data into an acoustic signal". Sonification is the core element of the "auditory display", which comprises "any technical solution for gathering, processing, and computing necessary to obtain sound in response to data" [1]. A comparison between visualization and sonification [2] can partially exculpate the much lower use of sonification for scientific purposes as compared to visualization. Nevertheless, the present technical advent and the permanent increase of expectations to extend the ways we interact with the environment open the doors for new developments in this direction.

This paper aims to present an overview of the potential approaches for various applications, to describe authors' experience and explore in more detail the potential future applications for medical data representation.

Methods

The good survey of the sonification domain, until 2011, was done by Hermann et al [1] to which one can add the news from the site of the International Community of Auditory Display [3].

There is a large variety of methods, but almost all start from the same basic principle: to associate the major physical parameter of a sound — frequency (pitch) — with a (major) property of the data to be represented. Other properties of the data set can be linked to other properties of the sound sequence – duration or intensity or even raising the complexity of representation by introducing timbre, rhythm, harmony (multiple sound channels), etc.

We will present here our approach [4], which has multiple similarities with the one in [1], starting with temporal sequences.

Formalization Levels

We will first refer to two major parameters of sound – frequency (pitch) and duration.

Frequency

From a physical point of view, there are two distinct levels, corresponding to the sonic output: continuous or discrete frequency spectrum. However, for the purpose of potential applications we will prefer to define three levels:

1. *Acoustic level* – with a continuous spectrum,
2. *Sonic level* (S) – with a discrete spectrum, having values belonging to the musical scale, and
3. *Musical level* (M) – a complex level, with multiple channels, rhythm, and harmony.

As the original set of data can have values expressed by numbers in any region, they are usually normalized (y_i), to fit an interval – the simplest $[0, 1]$ interval. A reference frequency (f_0) is also needed to yield results within the audible interval. Some practical values would be 440 Hz (A4 on musical scale) or 262 Hz (C4). Since the natural sound scale is exponential, for the acoustic level the sound will have the frequency f_i :

$$f_i = f_0 \times 2^{(y_i)} \quad (1)$$

For the S level, the f_i value will be rounded to one of the values from the scale.

Duration

One can choose to display sound for data originating from time series with a duration equal to the real duration of the corresponding event. However, most of the times the data are discrete; even continuous processes are sampled. Hence, we need a conventional rule for the transition between two consecutive sounds. There are two major possible transitions:

1. Continuous transition (also called sublevel A), when for two consecutive points, at t_i and t_{i+1} , the frequency will vary continuously from f_i to f_{i+1} ; and
2. Discrete representation, when frequency f_i will be displayed for the interval $dt = (t_i, t_{i+1})$, followed then by f_{i+1} . Usually the intervals dt are very short and this sub-level is called “quasicontinuous” (Q).

We have used the subdivision into A and Q sublevels only for the acoustic level, using the S level long durations, to be perceived as separate sounds.

There are several cases when the structure of the input data can suggest various corresponding relations between data characteristics and sound characteristics (often called ‘natural’ correspondence). Some such examples will be given in the section on potential medical applications.

Sound display

The large diversity of mapping data over sound parameters is still increased when thinking about the possibilities to display the results.

Tempolenses

For the cases when the duration of the analysed phenomenon is either too long or very short, and the duration of the sonic display is different from the recording, we can use compression or dilation procedures, also called “tempolenses” [5]. The main parameter of a tempolens (TL) would be the “magnification”, defined as the ratio between the sound display duration (t-repr) and the corresponding real process duration (t-real):

$$m = (t\text{-repr}) / (t\text{-real}) \quad (2)$$

When $m > 1$ the TL dilates the signal, good for detecting details (especially for very short events); when $m < 1$, it compresses the signal – good for exploring long recordings in a short time.

An interesting case is represented by TLs with variable magnification (TL-v). Unlike the TL with fix magnification (f), for TL-v, the magnification is $m > 1$ in certain regions and $m < 1$ in other regions – a version recommended for detailed representation of fast process, without an overall sacrifice of the total displaying time; for instance, for the ECG sonification, the QRS complex is a fast event and can be dilated, while the T–P interval does rarely contain relevant events and can be compressed. A tempolens with a variable magnification is like a “distorsion” tool, with potential applications.

Sound artifacts

Depending on the potential use, the conventionality degree can be increased by introducing various artifacts, with specific significance. This would be especially useful in monitoring systems associated with warnings. Two such artifacts are described here.

- *Saccadic display*

The auditory system contains phasic receptors, hence it has an efficient adaptation (decrease of the response at a constant value of stimulus intensity), hence an interrupted sound would be better perceived than a continuous sound. Thus, instead of a single sound, we can introduce a “saccad” of two or three sounds as a warning. Moreover, some patterns can be used for different situations; these patterns can easily be learned (when not too many of them).

- *Intensity (loudness)*

For different levels of warning, the intensity of the sound can also be varied, especially when reaching “alert” regions of the followed parameter (for example, heart rate during exercise).

Differential display

One of the potential future developments might rely on the capacity of the ear to distinguish close frequencies. We can compute an average normalized period for periodic signals, then display it repeatedly on one channel while the real signal is displayed on a second channel so that even small deviations will be perceived very easily.

Non temporal sequences

A relation similar to (1) for frequencies can be used for non temporal sequences such as the primary structure of nucleic acids or proteins, but the (y_i) array is not an ordered array. One can introduce an “order relation” into the original data starting from some properties which might be relevant such as molecular weight, hydrophobicity index, or percent abundance in a certain structure. The correspondence map does not have to follow the ordered musical scale, it might also be linked by some other properties such as probabilities to find a certain note in a musical composition. One can imagine the tremendous variety of mapping systems and it is difficult to qualify/score them.

Transforming digital images into sounds

A serial transmission of an image is actually a signal which can be sonified by any of the procedures described above, and therefore, indeed, such transformations might be useful.

Results

As this paper is submitted as a “vision” paper, this section does not present only our own results [6], but also other applications, all are regarded as examples to be discussed. All examples refer only to applications in medicine and life sciences.

Sonification of cardiac signals

Cardiac signals, especially the electrocardiogram (ECG) are periodic signals, often used for diagnosis in cardiovascular diseases.

Heart rate analysis

Cardiac heart rate (HR) is one of the simplest signals to record, either from pulse oxymeter devices or from ECG. It was one of the first attempts to sonify a physiological signal done by Ballora [7] in which he tried to raise the complexity up to the musical level. Instead, we tried to keep closer to the original signal [4], using a couple of different sets of sonification parameters and compared the results. Thus, we realized the sound produced by the acoustic level A would not be attractive (sounds like whistling), the same for Q sublevel with very short durations. Sublevel Q with larger durations or level S seem appropriate for detecting deviations from normal (sinus) rhythm. For recorded signals (we have used RRs and ECGs from Physiobank [8]), a compressed sonic display ($m=1:6$ to $1:4$) seemed to be most appropriate.

ECG analysis

We had bigger expectations to get better results from ECG analysis. Indeed, interpreting the classical recording was not an easy task, some modifications were very small, it was important to analyse both amplitudes and intervals, one had to look over the recordings in several leads. What would be the most convenient sonification procedure to reveal all these aspects?

- *Integrated display*

We tried to apply several sonification parameter sets for the same signals [6, 9], including various types of tempo-lenses [5], and compare them (Figure1) using an integrated display – both visual and sonic; a library of various signals and their sonification was created [10].

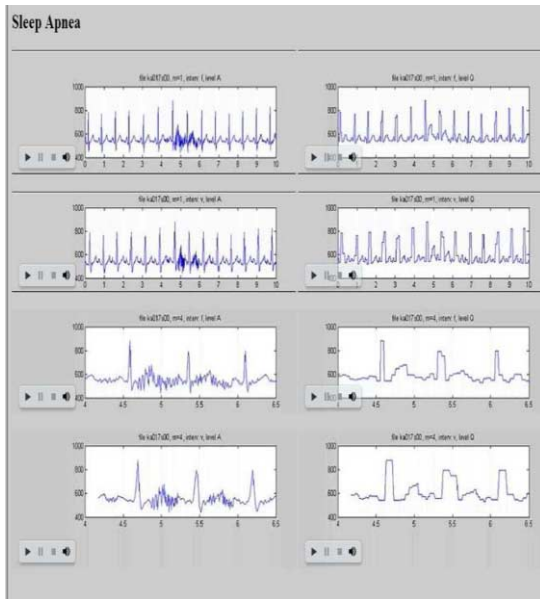


Figure 1 - Example of a screen for ECG signal sonification

- *Discriminant power*

An important issue in our project was to test the discriminant power of various sonic representations i.e. the capacity of listeners to distinguish details and recognize various types when the same signal is represented in different ways [11]. An example is presented in Figure 1: a 10 sec ECG signal was displayed in eight different forms (A or Q mode, f or v display, TL 1× or 4×). For Q mode we had $dt = 0.4$ sec. The sounds were presented in two groups, one of physicians (residents) and on of music students. We present here some relevant conclusions from this study:

- Representation in A mode was less preferred by both groups (it sounds like a whistle); however it had a higher discriminant power in sleep apnea, during the obstructive episodes;
- Tempolenses with variable magnification did not bring the expected increase in resolution of the QRS complex;
- Short durations (less than 0.2 sec) in Q mode sound like A;
- The distinction between Q and S modes was much clearer for the group of musicians; small differences for other cases.

Our studies showed that there were clear distinctions between the sonic display of different signals (normal sinus rhythm, arrhythmia, and atrial fibrillation), such a simple “additional sound” did not prove to be very attractive to physicians, since their classical procedures did fully satisfy their needs.

Cardiac parameters variation during exercise

We found an interesting application, with a higher user acceptance for monitoring cardiac parameters rate during exercise. Professional equipment can trace various parameters important for the exercise tests on cardiac patients, like the heart rate HR or depression of the ST segment in ECG recordings, with warnings when some parameters exceed thresholds. But this information is mostly visually displayed and the patient is usually kept passive.

We have developed an application to add sounds for various thresholds of HR [12] or ST segment [13]. The threshold warnings for HR have been established using Kevonen relations

[14], separating four exercise levels: quasi-rest/[start threshold]/aerobic mild exercise/[attention threshold]/hard exercise anaerobic zone/[alert threshold]/risk zone with $HR > HR_{max}$. Each level (zone) has a specific intensity, from very low to loud, and saccadic display (no saccades/1/2/3 saccades). The original record is trimmed, keeping the regions comprising transitions between zones.

An example is given in Figure 2 and the sonic displays can be downloaded from [10]. Three types of sonic displays have been tested: with various patterns of pitch/saccades/intensity. The preferences were oriented towards a warning display using four distinct patterns: no sound or low intensity continuous sound for normal (safe) domain, and increasing the number of saccades, the frequency and the intensity for each higher level.

The patients as well as the physicians monitoring the exercise tests appreciated the warning system.

An extension of the application for the self monitoring of the HR during daily individual exercise (jogging) has also been tested.

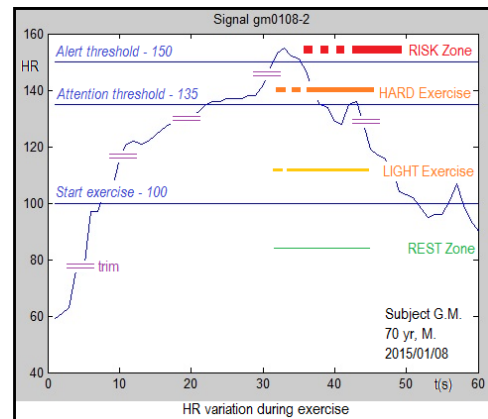


Figure 2 - HR zones during exercise and sound scheme

Molecular sequences

One of the most prolific areas of sonification was inspired by the similarity between musical sequences and biomolecular structures, especially DNA or proteins. Among early attempts we can cite Susumu Ohno's mapping of DNA [15], associating the nucleotides with notes of the musical scale (C cytosine to 'do/C', A adenine to 're/D or mi/E', G guanine to 'fa/F or sol/G' and T thymine to 'la/A or si/B') and applying some rules for durations linked to the repetitions in sequences. Several other algorithms have been developed: 'gene-2-music' project at UCLA (R Takahashi, JH Miller and F Pettit) [16], 'Algorithmic Arts' for protein sequences, by John Dunn [17], 'genetic-musicproject' by Greg Lukianoff [18].

We have also tried an algorithm based on the relations between physico-chemical properties of amino acids (hydrophobicity, polarity, size, and abundance) and the musical scale [10].

Ohno and Midori [15] have also tried the reverse correspondence – to transpose a musical piece (Chopin's Nocturne op.55) into a DNA sequence and compare the simulated molecular structure with real molecules, obtaining a certain resemblance with the structure of an enzyme (mouse DNA polymerase II).

There are now several sites (such as [17, 18]) allowing you to upload a sequence (real or imaginary), which generates a musical sequence. The results are attractive, but, despite the attempt to build algorithms based on 'natural' maps, often using statistical data processing or even more sophisticated procedures, the

impact on scientific community remains modest. This might be, at least partially, attributed to the low degree of usability.

Sonification of other types of data

Our experience include also other applications (not detailed here), like pulse wave signals in experiments on mice, cellular kinetics (especially protein-protein interaction for p53-mdm2 system), and molecular sequences (work in progress now).

We should mention here that there are several teams working in this field and we can cite several papers from the literature. A comprehensive site is the one of ICAD [3].

There are two other interesting applications in the biomedical domain:

1. Analysis of fetal heart rate, Paracelsus Clinic, Austria [19], and
2. Bio-rhythm analysis [20].

Discussion

The results presented above as well as a full list of references would rather give the impression, at a first glance, of a well set domain, with well defined tracks to be developed in the future. However, a thorough and fair analysis would reveal several weaknesses and inconsistencies that actually brought such a limited applicability in the bio-medical domain.

Analysis of the present state

We can point to the following as among the potential reasons to account for the reduced use of sonification:

- The technological support was not fully available for handling sounds. This point is no longer an issue.
- Most authors had tendency/ambition to reach the musical level. It seems attractive, indeed, but it introduces a high degree of arbitrary/conventional rules, moving us farther from the real signal. Not all real data carry harmony! Moreover, quite often the algorithms generated ‘atonal’ values that have been rounded to make them musical.
- A new tool would be accepted only if it brings something new, which is not easily achieved by present methods. Currently, visualization, complemented by the well-developed technical support, meets expert expectations in bio-medical practice.
- Even when a new method brings more information, the ones in use have several objective and subjective barriers.
- The limited usability is, perhaps, the most relevant factor; most of the work done in the sonification area started as an ‘interesting’ approach, almost as a curiosity, and not as a real need to solve a practical problem in a better (or cheaper or more convenient) way than by other methods.

Some such potential applications will be presented below.

Let’s learn from visualization

A comparison between the two methods might help us in finding better ways to strengthen the usefulness and use of sonification as a method for data representation.

Table 1– Comparison between Visualization and Sonification

Property	Visualization	Sonification
Best representation	spatial	temporal
Prevalent elements	static	dynamic
Type of attention	focused	distributed

Recommended for Detection of small difference	structures poor	processes good
Specific elements	brightness texture color shape	loudness harmony timbre rhythm

A deeper insight into this table will let us propose a set of properties of a sonic display, similar to that defined by Tufte for graphical/visual displays [21]; some requirements are valid for both types of displays (marked with *).

Thus, we can consider that a ‘sonic display’ should:

- (*) represent the data
- (*) avoid distorting what they do represent
- (*) make large data sets coherent
- (*) serve a clear purpose: description, exploration
- Preserve the initial sequence
- Bring out information that is inaccessible (regardless the reason) or unavailable by other means
- Be integrated with other informative displays, when possible.

The list above can (should) be refined for a more comprehensive and clearer meaning.

Potential bio-medical applications of sonification

The experience accumulated during the research—both positive and negative—would let us find the directions to be followed for future successful applications in the bio-medical domain.

One category of applications refers to the cases when the visual system is engaged in other activities, such as providing information during driving or jogging.

A second category would refer to all kinds of warnings, based on the property that sonic/acoustic warnings do not need a focused attention. It might become a useful tool in the context of the rapid development of wearable devices for monitoring various physiological parameters. We think this direction might turn into a prolific domain.

There are still less explored directions for applications to categories of people with attention disorders, like elderly people or children with ADHD.

Yet, one can also consider the potential exploration of large databases to find small differences between patterns or to detect mutations in molecular structures.

Last, but not least, we can also include several applications for visually impaired people – another direction which has not been explored enough.

This list is, of course, incomplete, and will have hopefully new items in the future.

Conclusion

Sonification is a method with several potential applications in medicine, yet it is not explored enough. The large palette of sonification methods, without well-defined procedures or standards and the quasi-random examples have unwittingly limited its application, generating a low user acceptance.

The discussion section, based on the present data and the previous work of the authors, tried to analyze the present state of the domain and anticipate future directions of work.

References

- [1] T. Hermann, A. Hunt, J.G. Neuhoff (eds.), *The Sonification Handbook*. Berlin: Logos-Verlag, 2011.
- [2] S. Pauletto and A. Hunt, A comparison of audio and visual analysis of complex time series data sets, *ICAD05* (2005), 175-181.
- [3] International Community for Auditory Display, Available at: www.icad.org (as on 04/12/2017).
- [4] G.I. Mihalas, S. Paralescu, N. Mirica, D. Muntean, M. Hancu, A. Tudor, and M. Andor, Sonic Representation of Information: Application for Heart Rate Analysis, *Quality of Life through Quality of Information. Proceedings of MIE2012* ID_403, Amsterdam: IOS Press, 2012.
- [5] G.I. Mihalas, S. Paralescu, M. Andor, and et al, Tempolenses with Variable Magnification for Sonic Representation of Medical Data. Application for Cardiac Signals, *Stud Health Technol Inform* **186** (2013), 78-82.
- [6] G.I. Mihalas, M. Andor, S. Paralescu, and et al, Adding Sound to Medical Data Representation, *Proc. 21st Intl Conf on Auditory Display ICAD 2015*: 325-326, 2015.
- [7] M. Ballora, B. Pennycook, P.C.H. Ivanov, and et al, Heart Rate Sonification: a New Approach to Medical Diagnosis, *Leonardo* **37**(1) (2004), 41-46.
- [8] A.L. Goldberger, L.A.N. Amaral, L. Glass, and et al, PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals, *Circulation* **101**(23) (2000), e215-e220.
- [9] G.I. Mihalas, S. Paralescu, A. Tudor, and M. Andor. Adding Sound to ECG, *Stud Health Tehnol Inform* **216** (2015), 1044.
- [10] Victor Babes University of Medicine and Pharmacy, Available at: <http://www.umft.ro/dim/sonification> (as on 04/12/2017).
- [11] M. Andor and et al, Testing Discriminant Power of Sonic Representation of ECG Signals, *Cercet. Exp. Medico-Chirurg* Available at: http://jmed.ro/articole_en.php?an=toti&id=365, 2013.
- [12] M. Andor, A. Tudor, S. Paralescu, and G.I. Mihalas, Methods for Sonic Representation of Heart Rate During Exercise, *Stud Health Tehnol Inform* **210** (2015), 60-64.
- [13] M. Andor, A. Tudor, S. Paralescu, and G.I. Mihalas, Methods for Sonic Representation of ST Depression During Exercise, *Stud Health Tehnol Inform* **216** (2015), 1041.
- [14] American Council of Exercise, Available at: www.acefitness.org/blog/1179/measuring-intensity, (as on 04/12/2017).
- [15] S. Ohno and M. Ohno, The all pervasive principle of repetitious recurrence. *Immunogenetics* **24** (1986), 71-78.
- [16] UCLA Microbiology, Immunology, & Molecular Genetics, Available at: http://www.mim.ucla.edu/faculty/miller_jh/gene2music/projectdevelopment.htm (as on 04/12/2017).
- [17] Genetic Music from DNA and Protein, Available at: <http://www.algoart.com/music.htm> (as on 04/12/2017).
- [18] Genetics Project Music, Available at: <http://www.geneticmusicproject.com> (as on 04/12/2017).
- [19] Heartmusic: Graphic & Audio Display of Maternal & Fetal HRV, Available at: www.michaelfalkner.de/herzmusik/heartmusicinfo.html
- [20] R. McGee, Y.Y. Fan, and R. Ali, BioRhythm: a Biologically-inspired Audio-Visual Installation, Available at: www.mat.ucsb.edu/Publications/McGee_NIME_2011.pdf (as on 04/12/2017).
- [21] E. Tufte, *The Visual Display of Quantitative Information*, Cheshire Connecticut: Graphics Press, 1983.

Address for correspondence

Corresponding author: George Mihalas, email at: mihalas@gmail.com.

Constructing a Gene-Drug-Adverse Reactions Network and Inferring Potential Gene-Adverse Reactions Associations Using a Text Mining Approach

Sui MingShuang, Cui Lei

School of Medical Informatics, China Medical University, Shenyang, Liaoning, China

Abstract

Our objective was to identify and extract gene-drug and drug-adverse drug reaction (ADR) relationships from different biomedical literature collections, and to predict the possible association between gene and ADR. The drug, ADR and gene entities were recognized by a CRF model with multiple features. Logistic regression models were constructed for each drug-ADR and drug-gene pair based on its frequency, Mesh Rule association and similarity with known association etc. Using predicted score to generate drug-ADR matrix and drug-gene matrix, and then calculating for gene-ADR matrix. Network and clustering analysis were applied to verify and interpret the relationship between them. A total of 78014 potential gene-ADR associations were predicted. Part of the predicted results can be explained by the network-clustering-pathway analysis, and verified in the literature. The gene-drug-ADR network constructed in this study can provide a reference for the possible association between the gene and ADR.

Keywords:

Drug-Related Side Effects and Adverse Reactions; Data Mining; Algorithms

Introduction

Adverse Drug Reactions (ADR) is a major burden on public health systems in the world, the complex mechanism of drug action and clinical trial makes pharmacovigilance long way to go. The onset of ADR is closely related to drug's interaction with its transporter, metabolizing enzyme and its targets, whose deletion and other changes can cause the corresponding genetic variation of expression products, further affected drug efficacy, or even serious ADR such as CYP2C9 and VKORC1 gene variants and warfarin caused bleeding [1]. Therefore, Detecting the relationship between drug related genes and ADRs can help to determine the specific response of individual gene to drugs, so as to improve drug safety and provide evidence for precision medicine.

Recently, a large number of text mining-based biomedical knowledge discovery researches have shown that text mining can be successfully applied to detect the relation between gene, protein, diseases and other biomedical entities [2, 3]. Measuring similarity between entities through structural pathway properties, intrinsic properties, and taxonomies can help to predict drug adverse reactions and targets[4, 5, 6]. For example,

Yamanishi constructed a high-dimensional matrix based on drug chemical structure, in which each drug was encoded according to whether specific drug features and the targets extracted from Pubchem was observed, and then predicted ADR using machine learning algorithms [7]. Liu applied chemical properties (substructures and fingerprints), biological properties (protein targets and pathways), phenotypic properties (indications and ADR) to predict potential ADR [8]. These studies, although different in approach, are confined to a single topic study sample. Professor Swanson (1985) proposed a knowledge discovery approach to discover valuable knowledge hidden in the medical literature based on non-related topics literature collection, which provides an insight into the predicting potential link between gene and adverse drug reactions in the literature [9].

This paper aims to recognize and extract gene-drug (D-G) and drug-ADR (D-ADR) associations from biomedical literature collections, and then infer potential association between genes and ADR by network and cluster results.

Methods

Overview

First, we integrated data from multiple databases, including literature collections, vocabularies and prior knowledge. After naming entity recognition, we extracted the D-ADR and D-G relationship by Logistic Regression (LR) model, thus forming hypothesis about G-ADR relationship. Ultimately we interpreted these inferred relationship by combining network analysis, clustering analysis and related literature information (see Figure 1).

Data Sets

Literature collections were obtained from PubMed database Abstract Section (search strategy A, ("Gene expression"[majr] OR "Gene expression regulation"[majr] OR Genes[majr] OR gene[ti] OR genes[ti]) AND "all terms in Pharmacological Actions Category"[Pharmacological Action] AND has-abstract[text] AND "humans" [MeSHTerms] AND "2006/11/30" [PDat] : "2016/11/30" [PDat]" for D-G data sets and strategy B, "all terms in Pharmacological Actions Category" [Pharmacological Action] AND adverse effects[sh] AND chemically induced[sh] AND (hasabstract[text] AND "2006/11/30" [PDat]:"2016/11/30" [PDat] AND "humans"

[MeSH Terms]) for D-ADR data sets. We downloaded the MEDLINE format file, and extracted title and abstract as the research corpus.

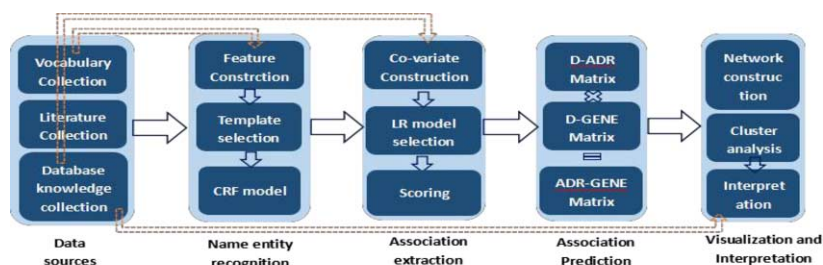


Figure 1- Method Overview

Drug vocabulary was obtained from Mesh and Drugbank [10], and ADR vocabulary was from Mesh and MedDRA (Medical Dictionary for Regulatory Activities), and gene vocabulary was from NCBI Gene database.

Prior knowledge was gained from CTD (Comparative Toxicogenomics Database) [11] for D-G associations, SIDER (Side Effect Resource) [12] for D-ADR associations, and Drugbank database for drug-targets associations. We mapped these associations to which identified by Mesh ID and NCBI Gene ID.

Name Entity Recognition and Relationship Extraction

After splitting abstracts into sentences, we applied a CRF-based model for Drug and ADR Name entity recognition in our previous work [14]. Gene recognition was generated from GNormPlus[15] and gene dictionary matching results.

We extracted the co-occurring D-ADR and D-G entities in a sentence as candidate pairs. To determine whether there is association between D-G (ADR) for each pair, we applied a knowledge-driven algorithm [16] to construct two LR models based on existing D-G (ADR) knowledge.

The covariates used in LR models are of two main types. Baseline covariates included D-G (ADR) co-occurrence frequency (F_c), key verbs frequency (F_v), Mesh Rule association (MR) (only for D-ADR), co-occurrence with prior knowledge in SIDER (F_s), and minimum distance (MINd). Taking a candidate D-ADR pair as an example, the key verbs refer to verbs that appeared between the drug and the ADR in a sentence, indicating that they have clear semantic relations, such as “cause”, “induce” etc[17]. In order to reduce the false positive rate (such as drug-treat-disease), we introduced MR covariate, it refers to semantic association in Mesh terms. For each pair of D-ADR, we checked out whether “drug/adverse effect” and “ADR/chemical induced” co-occurred in a article (Supplemental concept record would be mapped to the corresponding Mesh term according to “heading mapped to”). The MINd measured the minimum distance between the drug and the ADR in the sentence with key verbs between them.

The second covariate type was Network covariates, it depend on the structure of the prior D-ADR(G) network [18]. We constructed a bipartite network to represent the drugs, ADRs, and their associations in SIDER. In this network, nodes denote drugs or ADRs and edges denote known D-ADR associations.

These covariates aimed to capture the structural similarity between drug pairs and ADR pairs. Denoting variable i as a drug, variable j as an ADR, and ADR_i denotes the set of neighbors of node D_i , and D_j denotes the set of neighbors of node ADR_j .

1. JSMD(Jaccard-sider-drug-max): For each (D_i, ADR_j) , we introduced JSMD covariate to measure the max similarity between the D_i with ADR_j 's neighbor nodes D_j . It was computed as follow,

$$JSMD(i, j) = \max_{k \in ADR_j} \{J(j, k)\}$$

$$J(j, k) = \frac{ADR_j \cap ADR_{(k)}}{ADR_j \cup ADR_{(k)}}$$

2. JSAM(Jaccard-sider-adr-max): For each (D_i, ADR_j) , we measured the max similarity between the ADR_j with D_i 's neighbor nodes ADR_i .
3. JDDM(Jaccard-drugbank-drug-max): Studies have shown that drugs with a common target tend to lead to same or similar ADR [19]. Therefore, we constructed a drug-targets bipartite network based on drug-target information in Drugbank, and defined Target (i) as the neighbor nodes of the node D_i . For each pair of D_i - ADR_j , the maximum similarity of the ADR_j 's neighbor node set D_j with D_i of was calculated.
4. JCDM(Jaccard-ctd-drug-max): In order to support the hypothesis that the genes interacting with the same drug may affect its ADR, we constructed a bipartite network using the D-G association in CTD, and defined D_i 's neighbor node set Gene (i), and measured the similarity between drugs based on their interactive genes. We named it interactive network covariate.

Similarly, for D-G, we calculated drug similarity based on SIDER (Jaccard-sider-drug-max, JSMD), drug similarity based on Drugbank (Jaccard-drugbank-drug-max, JDDM), drug similarity based on CTD (Jaccard-ctd-drug-max, JCDM), gene similarity based on CTD (Jaccard-ctd-gene-max, JCGM).

After the multicollinearity diagnosis, the variables were imported into the LR model, and the optimal parameters were set according to the model results. The D-G matrix and the D-ADR matrix were generated. The G-ADR matrix was obtained by multiplying the above two matrices [20] as the potential basis for G-ADR prediction.

Network and clustering analysis

Network analysis can help us to understand the resulting matrix and association, we constructed three types network according to D-G matrix, D-ADR matrix with Cytoscape [21] (for the global network) and Gephi [22] (for local network). Additional, we carried out systematic clustering analysis on the G-ADR matrices. The cluster results were expressed according to the principle of TF-IDF [23].

Results

The corpus used in study contained 59,660 and 64,499 abstracts from MEDLINE for D-G and D-ADR literature collection, corresponding to 63,259 and 114,919 co-occurrence pairs respectively.

Relationship Extraction Results

LR analysis showed that F_c , MR, MINd, F_s , JSDM, JSAM, JDDM, JCDM covariates were included in D-ADR model, and F_c , F_{cid} (co-occurrence frequency with CTD), MINd, JSDM, JCAM and JDDM in D-G model, with AUROC 0.986 and 0.952 respectively.

The two LR models both contain the interactive network covariate JCDM and JSDM, which indicated that gene (ADR) - based similarity is meaningful variables for predicting D-ADR (G), which provided an indirect explanation for our study. Based on predictive scores(cutoff 0.5), a D-ADR matrix (854 * 717), a D-G matrix (764 * 2,153) and then a G-ADR matrix (1,488 * 665) was generated.

Network analysis

We constructed D-ADR and D-G networks by Cytoscape and Gephi. The sub-network structure was shown in Figure 2, the left orange node represents the gene, the middle green node represents the drug, and the right purple node represents the ADR. The two color edges indicate the D-G and D-ADR relationships (using predicted scores as weight). It can be seen that 17 kinds of genes and 29 kinds of ADRs were linked by tamoxifen, corresponding to 442 potential G-ADR associations. For the G-ADR association, the more drug nodes connected to each G-ADR pair may indicate a more potential association between these drugs, which can be inferred that the gene may be associated with ADR of certain drugs. Compared with the results of G-ADR matrix, network analysis can provide information on the drug medications.

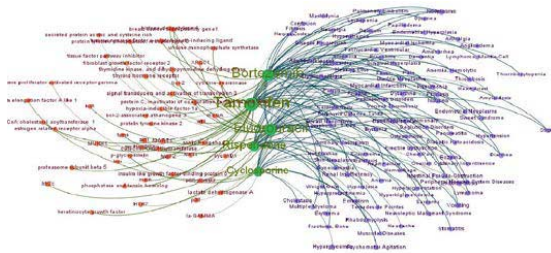


Figure 2- Sub-network structure

We selected Drug Hypersensitivity Reactions as a case study. Drug Hypersensitivity Reactions is a serious systemic drug response, mainly characterized by acute extensive lesions, with fever, lymphadenopathy, multiple organ involvement (hepatitis, nephritis, pneumonia), eosinophilia and mononuclear cells and other hematologic abnormalities, including Eurasian disease, Drug Eruptions, Hand-Foot Syndrome, Erythema Nodosum, Acute Generalized Exanthematous Pustulosis, Stevens-Johnson Syndrome and other ADR. Network analysis showed that these ADRs could be associated with 857 genes through 179 drugs. A total of 1,713 gene-ADR pathways are related via different drugs. Among them, some associations have been confirmed in the relevant studies, such as hypersensitivity can be associated with and Human Leukocyte Antigen (HLA) [24] through Carbamazepine as shown in Figure 3. In addition to HLA, the network can also infer unc-13 homolog B, microsomal epoxide hydrolase, uchl1, monoamine oxidase B, voltage-gated sodium channel α subunit type I, dopamine receptor D2, tachykinin 3 and amphiregulin may be related to the onset of hypersensitivity induced by drug.

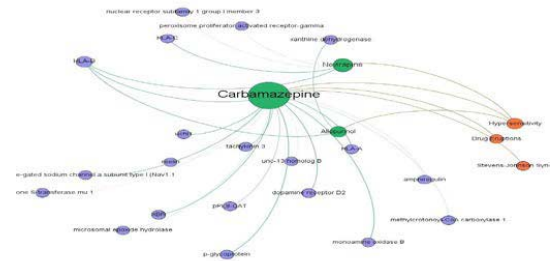


Figure 3- Drug Hypersensitivity Reactions related network

Network clustering analysis results

Clustering analysis can help better understand for ADR and gene mechanism with drug and allow us to grasp the potential association from the perspective of the overall cluster. For the integrated ADR-D-G network, we use the Cytoscape plug-in Auto Annotate community detection algorithm to obtain the corresponding clusters and interpret them with the cluster topic. As shown in Figure 4, the green, blue and yellow nodes represent drugs, ADRs and genes respectively. These nodes can be divided into 10 clusters and several subgroups. There are also associations between clusters.

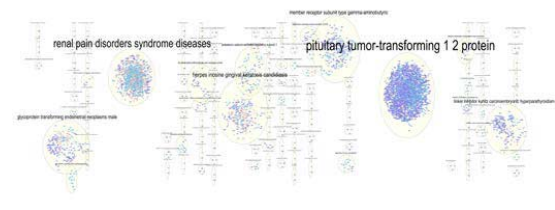


Figure 4- Results for network clustering analysis

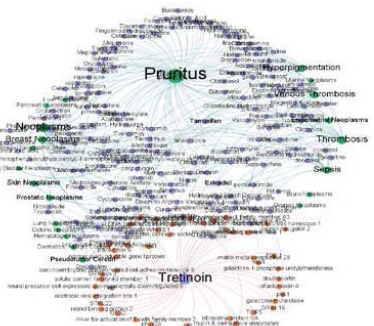
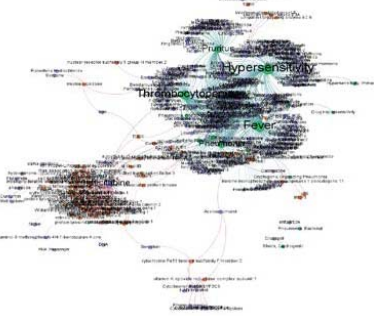

Matrix clustering analysis

We carried out clustering analysis for the G-ADR matrix and obtained representative ADR that can explain the content for each cluster. In order to facilitate understanding, we selected several serious ADR to explain the three associations combined with REACTOME pathway information, the G-ADR clustering results are shown in Table 1.

CLUSTER 1: The results of clustering analysis combined pathway information showed that these genes mainly involved in

retinoic acid pathway, signal transduction pathway, immune metabolism pathway, and have potential association with skin lesions related ADRs represented by pruritus, rash, pigmentation. Signal transduction genes, such as retinol-binding proteins, are vitamin A transporters. Vitamin A plays an important role in maintaining normal physiological function of epithelial cells and promote the synthesis of immunoglobulin. More than normal values will lead to dry skin, scaling and hair loss and other symptoms, can also cause pseudo-brain tumor; Vitamin A acid as the metabolic

Table 1- Results for G-ADR Clustering analysis

cluster	Genes with pathway information	Representative ADR	Related Network
1	<i>Retinoic acid pathway</i> : aldehyde dehydrogenase 1 family member A3, retinol dehydrogenase 10 (all-trans); <i>Signal transduction pathway</i> : retinol binding protein 2; <i>Immune metabolism pathway</i> : linker for activation of T-cells family member 2, STRL22; <i>Others</i> : solute carrier family 40 member 1	Pruritus Hyperpigmentation Pseudotumor Cerebri Dermatitis	
2	<i>Vitamin K metabolic pathways</i> : vitamin K epoxide reductase complex subunit 1 <i>Hemostatic pathway</i> : myeloproliferative leukemia virus oncogene, integrin subunit alpha, solute carrier family 8 member A21 <i>Platelet-associated pathways</i> : myeloproliferative leukemia virus oncogene Wnt <i>Neutrophil degranulation</i> : myeloperoxidase, Rho GTPase activating protein 45	Thrombocytopenia Neutropenia Pneumonia Hypokalemia Febrile Neutropenia Pruritus Hypersensitivity	
3	<i>Zinc finger protein</i> : solute carrier family <i>Cholesterol synthesis pathway</i> : seladin-1, ISLOS <i>Metal ion transport pathway</i> : potassium voltage-gated channel subfamily J member 11	Pancreatitis Rhabdomyolysis Muscular Diseases Flushing Myalgia	

intermediates of vitamin A, can regulate the melanin produced by melanoma cells, pigmentation may be due to drug-effected Vitamin A acid pathway. Other gene such as solute carrier family 40 member 1 is directly involved in the pathogenesis of hemochromatosis. Combined with the local network, these genes and ADR is the main mediator of tretinoin.

CLUSTER 2: It was showed that genes involved vitamin K metabolic pathways, hemostatic pathway and platelet-associated pathways may be involved in the development of thrombocytopenia. Immune-related pathways may be responsible for pneumonia, rash, neutropenia and other ADR. The mechanism of hypokalemia is associated with cisplatin and diazoxide, which

may affect the potassium ion-gated channel. Network analysis showed that integrin subunit alpha 2b was associated with thrombocytopenia via aspirin, decitabine and so on. Decitabine mediates the association of thrombocytopenia, pneumonia, febrile neutropenia and numerous genes such as interferon beta 1, thrombomodulin and so on.

CLUSTER 3: Topic of this cluster is clear and representative ADR was pancreatitis. Pathway analysis revealed that the major difference between the cluster and others was zinc finger protein and metal ion transport pathway related gene. Network analysis showed that pancreatitis was linked to these genes through Zinc, hydrochlorothiazide and other drugs, most of which related to metabolic pathway with zinc, such as zinc finger protein. In addition, potassium voltage-gated channel subfamily J member 11 is involved in energy metabolism and muscle formation, which may be helpful to explain the ADR of rhabdomyolysis.

Discussion

Here, we proposed a pattern for generating hypothesis for G-ADR relationship by integrating D-G and D-ADR association extracted from two different literature collections. Petric I had applied similar method for identification of relations between biomedical concepts in disconnected sets of articles [25]. Finally, we constructed a G-D-ADR network by integrating knowledge of databases, and detected the potential association between gene with ADR. Our study proved the possibility of knowledge discovery from large data for basic and clinically relevant research and will lay the foundation for this field.

The limitations of this study are mainly three aspects, the first was many variables built for relationship extraction in this study based on SIDER/CTD knowledge, which was confirmed good predictive performance for the new D-ADR/G relationship [18], but this method was disadvantageous for drugs not in database (Because they are not included in the prior knowledge network); Secondly, there should be greater efforts to filter G-ADR association we inferred by weakening genes/proteins weights that are related to most drugs, such as p-glycoprotein, those unspecific gene may lead to a higher false positive rate. Lastly, the goal of our study was to demonstrate the utility of the integrated network, not to develop an optimal model which is a broader objective.

Our D-G associations allows multiple further analysis, such as clustering analysis and pathway analysis. By clustering analysis, we observed that some ADRs-Genes association could be partly explained by pathways such as Stevens- Johnson and HLA-B [24]. In addition, a lot of new G-ADR associations have been found through the calculation of network link prediction, and some of them are currently unexplainable. These associations may be false-positive or may not yet be found. It is also possible that the gene and drug ADR interact through a cascade of reactions, not just through a gene or protein. We recommend that more attention should be devoted to severe ADR.

Conclusion

Our proposed approach for integrating D-G and D-ADR associations is intended as a complementary hypothesis generation tool to identify potential G-ADR relationships. Further research directions will focus on defining the relationship between drug and gene types, and integrate more complex biological pathway information to provide more accurate evidence for D-G prediction.

References

- [1] Tomek A1, Mařořka V, Kolářová T, et al. The bleeding risk during warfarin therapy is associated with the number of variant alleles of CYP2C9 and VKORC1 genes. *Cardiology*. 2013;125(3):182-91.
- [2] Karimi S, Wang C, Jimenez A M, et al. Text and data mining techniques in adverse drug reaction detection[J]. *ACM Computing Surveys*, 2015, 47(4): 56.
- [3] Tatonetti NP, Liu T, Altman RB. Predicting drug side effects by chemical systems biology[J]. *Genome Biol*. 2009;10(9):238.
- [4] Campillos M, Kuhn M, Gavin AC, et al. Drug target identification using side-effect similarity[J]. *Science*. 2008, 321(5886):263-6.
- [5] Fukuzaki M, Seki M, Kashima H, et al. Side Effect Prediction Using Co-operative Pathways[C]// IEEE International Conference on Bioinformatics and Biomedicine. IEEE, 2009:142-147.
- [6] Brouwers L, Iskar M, Zeller G, van Noort V, Bork P: Network neighbors of drug targets contribute to drug side- effect similarity[J]. *PLoS One* 2011, 6(7):e22187.
- [7] Yamanishi Y, Pauwels E, Kotera M. Drug side-effect prediction based on the integration of chemical and biological spaces[J]. *J Chem Inf Model*. 2012, 52(12):3284-92.
- [8] Huang LC, Wu X, Chen JY. Predicting adverse side effects of drugs[J]. *BMC Genomics*. 2011, 23(12)Suppl5:S11.
- [9] Swanson D R. Fish Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge[J]. *Perspectives in Biology & Medicine*, 1986, 30(1):7-18.
- [10] Law V, Knox C, Djoumbou Y, et al. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. 2014, 42(Database issue):D1091-7.
- [11] Mattingly C J, Colby G T, Forrest J N, et al. The Comparative Toxicogenomics Database (CTD).[J]. *Environmental Health Perspectives*, 2003, 111(6):793-5.
- [12] Kuhn M, Letunic I, Jensen LJ, et al. The SIDER database of drugs and side effects. *Nucleic Acids Res*. 2015 Oct 19.
- [13] Lafferty J D, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models For Segmenting And Labeling Sequence Data[C]// 2001:282–289.
- [14] Sui MS, Cui L. A CRF Model Combining Multiple Features for Chemical - Disease Named Entity Recognition[J]. *Data Analysis and Knowledge Discovery*, 2016, 32 (10): 91-97.
- [15] Wei C H, Kao H Y, Lu Z. GNormPlus: An Integrative Approach for Tagging Genes, Gene Families, and Protein Domains.[J]. *Biomed Research International*, 2015, 2015.
- [16] Xu R, Wang Q. Comparing a knowledge-driven approach to a supervised machine learning approach in large-scale extraction of drug-side effect relationships from free-text biomedical literature.[J]. *BMC Bioinformatics*, 2015, 16(5):1- 8.
- [17] Wang XY. Semanti Relations Extraction Based on MeSH term Association Rule--A Case Study of Drug Side Effects Causing Disease[D]. China Medical University, 2012.
- [18] Cami A, Arnold A, Manzi S, et al. Predicting Adverse Drug Events Using Pharmacological Network Models[J]. *Science Translational Medicine*, 2011, 3(114):3371-3378.
- [19] Brouwers L, Iskar M, Zeller G, et al. Network Neighbors of Drug Targets Contribute to Drug Side-Effect Similarity[J]. *Plos One*, 2011, 6(7):e22187-e22187.
- [20] Oki N O, Edwards S W. An integrative data mining approach to identifying adverse outcome pathway signatures[J]. *Toxicology*, 2016, s350–352:49-61.
- [21] Shannon P, Markiel A, Ozier O, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks[J]. *Genome Research*, 2003, 13(11):2498-504.
- [22] Jacomy M, Bastian M, Heymann S. Gephi: An Open Source Software for Exploring and Manipulating Networks[C]// International Conference on Weblogs and Social Media, Icwsm 2009, San Jose, California, Usa, May. 2009.
- [23] Cui L, Sui MS. Study on an Approach to Presenting the Co-word Clustering Analysis Results[J]. *Journal of the China Society for Scientific and Technical Information*, 2015, 34(12):1270-1277.
- [24] Tangamornsuksan W, Chaiyakunapruk N, Somkrura R, et al. Relationship between the HLA-B*1502 allele and carbamazepine-induced Stevens-Johnson syndrome and toxic epidermal necrolysis: a systematic review and meta-analysis. *JAMA Dermatol*. 2013, 149(9):1025-32.
- [25] Petric I, Urbancic T, Cestnik B, et al. Literature mining method RaJoLink for uncovering relations between biomedical concepts[J]. *Journal of Bio-medical Informatics*, 2009, 42(2):219-227.

Address for correspondence

Cui Lei, E-mail:lcui@mail.cmu.edu.cn

Establishment of New National Rare Disease (Nambyo) Registry and Registry Guidelines in Japan

Hiroshi Mizushima, Mai Tanabe, Yasutaka Sugamori, Yoko Sato, Hiromitsu Ogata

National Institute of Public Health, Saitama, JAPAN

Abstract

A New legal structure for rare disease (nambyo) has been established in Japan this year, after 42 years of measures of nambyo. We have been accumulating registry for nambyo from 2003, however, as it was based on paper registration, quality was not enough.

Our new registry system will be based under ISO13606 which is a new medical international standard. Authorized doctors can put in data On Line by the new system, which has data cleaning filter for accurate data entry. Patients will be supported their medical expense by authorization by this system, so the registry will be efficient.

Keywords:

Rare Diseases; Registries; Demography

Introduction

In 1972, the Ministry of Health and Welfare enacted the "General Outlines for taking Measures to deal with Intractable Disease [NAMBYO]" which was the first one ever established in the world. The extent of Nambyo was proposed as : a) Unknown etiology b) Un-established therapy c) Frequent after-effects d) Chronic course e) Heavy burdens in terms of economical, psychological and physical issues.

At this stage, there was no concept of rareness in the extent of Nambyo. More than 25 years was needed until rareness was added to the extent.

A revised definition of Nambyo in 1995 was: a) Etiology is unknown. b) Frequency is low (less than 50,000). c) Therapy is not established. d) Economical, psychological and physical burdens. e) Diagnostic criteria should be established.

Frequency in Japan (< 4/10,000 : below 50K) is similar to that in EU (< 5/10,000) or in US (< 7/10,000: below 200K)

Selection of Nambyo

One hundred and thirty diseases were selected as Nambyo for special research. Two hundred and fourteen diseases were further selected as candidates of the second group of Nambyo for research support. Of 130 diseases, 56 diseases were targets for being treated with special favor in terms of medical expenses (see Table 1).

Budgets for researches (for 364 diseases) are 10 billion yen, and budgets for medical expenses (for 56 diseases) are 35 billion yen. Number of recipients of medical expenses are approximately 700,000 patients and increased by 30,000/year.

Nambyo was reviewed and a new legal structure was established in June 2014. This included: medical system, subsidized health care costs, promotion of research, consultation and support, human services, work and

employment support, international cooperation, and awareness.

At this time, supporting diseases were expanded from 56 to about 300, which will be defined by the committee and authorized by the government. This will double the patient numbers. New law will be effective from January 1, 2015. There are 330 from April 2017.

However, the most important issue to date is that the current registry system is done in paper form and not well organized.

As a result, we established a new registration system, which doctors input the data themselves which is then sent to an online national database.

This will be the first national online database for rare disease, and will support the patient care and research for rare diseases.

Methods

Systematic design of the registration system was done by analyzing the current paper-based registry system. Research in international standardization of terminology and modeling architecture was also completed. Prototype system was constructed with Mitsubishi Space Software, NISplus, and R102 company.

Results

Problems of Nambyo

Comprehensive measures for Nambyo are important, however demographics of the 56 diseases are changing. As shown in Figure 1, some diseases, such as Ulcerative Colitis, Parkinson's Disease and SLE were included from the beginning, but they already exceed the patient size (50,000).

Also, as this registry is based on the application for the medical expense reimbursement, doctors tend to write more severe notes for better approval by the government.

New Registry System

A diagram for new registry system is shown in Figure 2. Any doctor can complete the form in the current system, however, only certified doctors for Nambyo can write (input) the form. This will make the report more accurate. Data will be inputted in to the system, and checking will filter inaccurate data entries. At this point, the doctors can be notified when they have the patient data (in current system, it required long time as government staff asks the doctor after the application).

Also, it was prefectural government's task to input the paper allocation to the database system, and not required for the patient reimbursement. As a result, some prefecture did not get inputted in to the national registry. Doctor's data entry is

nessesary to make the application for the prefectural government, to ensure all patient information should be in the system comprehensively.

International Standardization

International collaboration is very important in rare diseases, as the patient size is limited for each disease. There are many standard coding systems, however each one varies per country. We have searched for many registries and many international discussions about which terminology to use, and which clinical modeling should be used.

As rare diseases are premature in clinical research, many diseases are not well described in the terminology. Even in ICD-10, most of the rare diseases in Orphanet (<http://orpha.net/>), the most well known rare disease directory) are not coded. This is also the case for Japanese standard nomenclature, SNOMED, and HL7.

We are still waiting for this international standardization, but currently following Orphanet as disease names, CIMI for clinical information Modeling, and ISO13606 for archetype.

As for the minimum dataset for the registry, we have compared it with Global Rare Disease Registry (GRDR <http://rarediseases.info.nih.gov/>) at NIH/NCATS, and EUCERD/Epirare (<http://eucured.eu>)

Grouping Datasets as Modules

Current registry form is defined by each disease, and no there was no concept of systematic cross disease analysis. There were difference in units or abbreviations between diseases. As we are designing a new registry system, we wanted to standardize the data set, and also wanted to be systematic as we have to add about 250 new diseases.

Figure 3 currently shows 14 Neurological Nambyo grouped into 4 groups and extracts common data sets for each groups. From this analysis, we were able to define a dataset as set module so we can design a new registry by just combining the modules.

We also made the dataset module management software, to make the registry for new diseases a systematic process.

Web-Based Entry System

Our new registry looks like Figure 4. Sections are categorized by tabs which have a) Basic information Diagnosis; b) Onset and Prognosis; c) Clinical evidence; d) Laboratory findings; e) Severity; f) Differential diagnosis; and g) Therapy and care.

There are filters and mandatory defined datasets. Tabs will be red if mandatory data is not sufficient or data is out of range.

Offline Registration Tool

There are some hospitals where internet access is restricted. Some local governments also have restrictions in internet use. To support data entry in these conditions, offline data entry application were developed. Data entry is similar with the on-line system however, data will be stored locally with encryption, and will be uploaded with other methods. Printing application forms for local government can be performed using this application.

This will also enhance the security issue using internet, as data upload is much more safer than online data entry.

Discussion

We are planning to perform a pilot study using this system at some hospitals along with local government installation.

Total numbers of registry entries will be about 1,500,000 every year. This will be one of the largest registry for rare disease in patient numbers.

This will be the first National Registry System by remote data entry with standardized coding. We are hoping to include other registries in the near future.

Conclusion

A New legal structure for rare disease (Nambyo) has been established in Japan this year, after 42 years of measures of Nambyo. We have been accumulating registry for Nambyo since 2003, however, as it was based on paper registration, quality was not strong enough.

Our new registry system will be based under ISO13606, which is a new medical international standard. Authorized doctors can input data online by the new system, which has a data cleaning filter for accurate data entry. From the discussion about integrating the data from registry for each diseases, we established a guideline in Japan. This guideline is now in the process of finalization and authorization.

01 Behcet disease	Parkinson disease	Shetler syndrome - fatal	56 Diarrheocephalo-hydropathic
02 multiple sclerosis	21 amyloidosis	familial incontinentia	dysfunction - Syndrome of
03 myasthenia gravis 0	22 ossification of posterior	39 primary pulmonary hypertension	abnormal secretion of
4 systemic lupus erythematosus	longitudinal ligament (CPL)	40 neurofibromatosis type 1,	prolactin. Syndrome of
(SLE)	23 Huntington disease	neurofibromatosis type 2	abnormal secretion of
05 subacute myelopathy	24 moyamoya disease	41 subacute sclerosing	gonadotropin. Syndrome of
(SMON)	25 Wegener granulomatosis	panencephalitis (SSPE)	antidiabetic hormone -
06 aplastic anemia	26 dilated cardiomyopathy,	42 Budd-Chiari syndrome	abnormal secretion of
07 sarcoidosis	congenital cardiomyopathy	43 idiopathic chronic pulmonary	thyroid
08 amyotrophic lateral sclerosis	27 multiple system atrophy	thromboembolism with	stimulating hormone -
(ALS)	ataxia telangiectasia	pulmonary hypertension	Cushing disease
09 scleroderma - dermatomyositis,	(SMD)	44 lysosomal storage diseases -	Acromegaly - Hypophosphatemia
or polymyositis	choreoacanthocytosis	fatty disease - other	
10 idiopathic thrombocytopenic	atrophy (CPA) - Shy-Drager	lysosomal storage diseases	
purpura	syndrome	45 adrenoleukodystrophy (ALD)	
11 (1) polyarteritis nodosa (2)	28 epidermolysis bullosa	46 Familial hypercholesterolemia (
microscopic polyangiitis	29 psoriasis	homocystinuria type)	
12 ulcerative colitis	30 disseminated spinal canal	47 spinal muscular atrophy (SMA)	
13 Takayasu arteritis	stenosis	48 spinal and bulbar muscular	
14 thrombocytopenic purpura,	31 primary biliary cirrhosis	atrophy (SMA)	
Burger disease	32 severe acute pancreatitis	49 Chronic inflammatory	
15 periphallitis	33 idiopathic necrosis of the femoral	dermatomyositis	
16 spinocerebellar degeneration	head	Polymyositis	
17 Crohn disease	34 mixed connective tissue disease	50 Hypertrophic cardiomyopathy	
18 fulminant hepatitis	35 primary immunodeficiency	51 Restrictive cardiomyopathy	
19 malignant rheumatoid arthritis	syndrome	52 Mitochondrial disease	
(Rheumatoid vasculitis)	36 idiopathic interstitial pneumonia	53 lymphangioleiomyomatosis	
20 Parkinson disease and related	37 retinitis pigmentosa	54 Severe erythema multiforme	
diseases - progressive	38 prion diseases - Creutzfeldt-	(Acute Phase)	
supranuclear palsy -	Jakob disease (CJD) -	55 Ossification of the ligamentum	
corticobasal degeneration -	Gerstmann-Sträussler-	flavum	

Table 1 – List of 56 diseases of Nambyo which Patients are Supported for Their Medical Expense

The movement of Number of Recipients of Medical Expenses

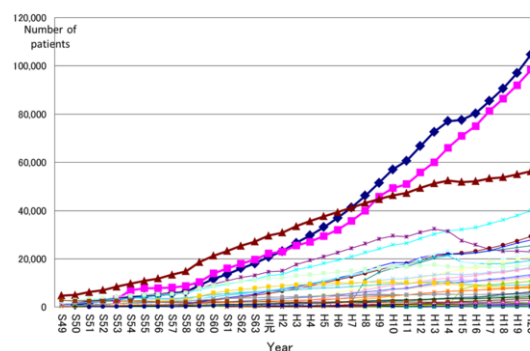


Figure 1 – Number of Patients from 1974 to 2008

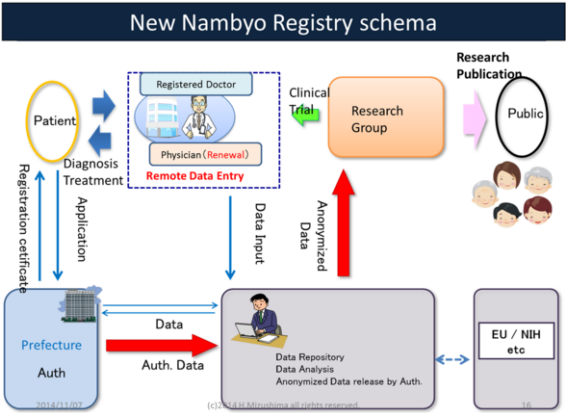


Figure 2 – New Nambyo Registry Schematic Diagram

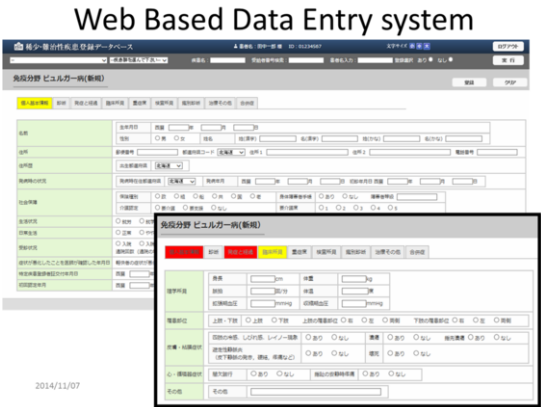


Figure 4 – Web based Data Entry System

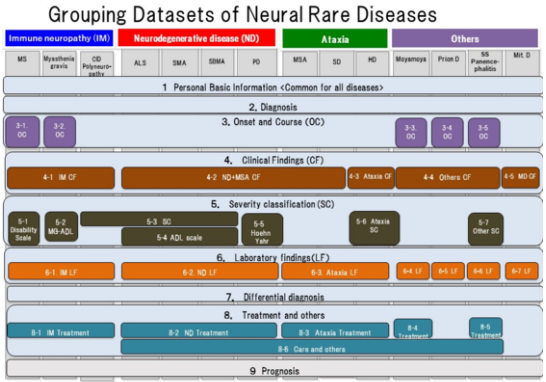


Figure 3 – Grouping of Datasets for Neurological Diseases

References

- [1] H.Mizushima, M.Tanabe, Y.Sato, Y.Kanatani; New legal structure and National Registry system for Rare disease (Nambyo) in JAPAN. IRDiRC conference 2014 Abstract book (2014).
- [2] OrphaNet. The portal for rare diseases and orphan drugs. Retrieved from <http://www.orpha.net/>
- [3] National Institute of Health. Genetic and rare disease information center. GARD. Retrieved from <http://rarediseases.info.nih.gov/>
- [4] European Union Community of Experts on Rare Disease. EUCERD. Retrieved from <http://www.eucerd.eu>

Address for correspondence

Hiroshi Mizushima
hmizushi@niph.go.jp
Tel/Fax: +81-48-458-6164 2-3-6
Minami Wako, Saitama 351-0197 JAPAN

Unsupervised Abbreviation Expansion in Clinical Narratives

Michel Oleynik^a, Markus Kreuzthaler^{a,b}, Stefan Schulz^a

^a Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Austria

^b CBmed GmbH, Center for Biomarker Research in Medicine, Graz, Austria

Abstract

Clinical narratives are typically produced under time pressure, which incites the use of abbreviations and acronyms. To expand such short forms in a correct way eases text comprehension and further semantic processing. We propose a completely unsupervised and data-driven algorithm for the resolution of non-lexicalised and potentially ambiguous abbreviations. Based on the lookup of word bigrams and unigrams extracted from a corpus of 30,000 pseudonymised cardiology reports in German, our method achieved an F₁ score of 0.91, evaluated with a test set of 200 text excerpts. The results are statistically significantly better ($p < 0.001$) than a baseline approach and show that a simple and domain-independent strategy may be enough to resolve abbreviations when a large corpus of similar texts is available. Further work is needed to combine this strategy with sentence and abbreviation detection modules, to adapt it to acronym resolution and to evaluate it with different datasets.

Keywords:

Natural Language Processing; Electronic Health Records.

Introduction

Clinical narratives are typically produced under time pressure. Besides incomplete sentences and typing errors, a typical manifestation of this is the tendency towards short forms like acronyms and abbreviations [1–3]. Their widespread use complicates further semantic processing and makes text understanding difficult, not only by laypersons, but also by clinicians from different disciplines or professional groups. In the context of large-scale concept mapping of medical texts for secondary use [4], the first processing steps are commonly devoted to text cleansing. This involves the correction of misspellings, the identification of sentence delimiters, and the expansion of short forms.

The use of abbreviations is especially pronounced in agglutinative languages such as German, in which single words are often formed by composition of morphemes, normally stems and affixes. The longer such a word grows, the more easily its completion can be guessed. Typing under time pressure therefore tends to stop close to the ending and an abbreviation marker is set. Thus, so-called ad-hoc abbreviations are created. Interestingly, the trailing period, required by German grammar to mark abbreviations, is mostly set, even in non-standard clinical narratives. It is also important to note that German requires capitalisation of all nouns, not only proper names, a rule commonly followed in clinical notes.

In combination, these characteristics lead to several NLP challenges related to the ambiguity of the period mark, which can be: (a) the trailing character of an abbreviation; (b) a sentence delimiter; or (c) both. Sentence delimitation, tokenization, and abbreviation resolution are, therefore, hardly separable tasks.

Regarding abbreviation resolution, three specific tasks are commonly listed: abbreviation detection, sense detection, and sense disambiguation. While the first is related to the disambiguation of the period character, the second and third relates to the sense expansion and its disambiguation, respectively.

Our recent work has focused on the first task, *viz.* abbreviation detection, first with a supervised approach based on support vector machines [5] and then with an unsupervised strategy that combined co-occurrence information with a large abbreviation-free domain dictionary [6]. The current paper will focus on the expansion (second task) and disambiguation (third task) of so-called period abbreviations, *i.e.* tokens that include a period mark as their rightmost character. The proposed approach is completely unsupervised.

Period abbreviations are clearly distinguished from acronyms, which consist mostly of upper case letters, never end with a period, and often represent multiword terms (e.g. “MI” for “myocardial infarction”). They are also distinct from other abbreviations that do not end with a period (e.g. “Ca”, which stands both for “Cancer” and “Calcium”).

Period abbreviations have, therefore, the following characteristics: (i) they abbreviate mostly single words; (ii) their first character always coincides with the first character of the word they abbreviate; (iii) in most cases, the string left of the period equals a string of characters from the left side of the abbreviated word.

Moreover, we distinguish between lexicalized period abbreviations and ad-hoc period abbreviations. Whereas the left substring rule can normally be taken for granted for the latter ones, lexicalized period abbreviations can be easily expanded via a lexicon. We also distinguish between unambiguous and ambiguous period abbreviations. General and domain specific dictionaries often list more than one sense for an abbreviation, especially for very short ones. Longer, ad-hoc period abbreviations are normally unambiguous. Conversely, single letter abbreviations like “E.” in “E. coli” for “Escherichia coli” are almost always ambiguous.

The following work is limited to period abbreviations that cannot be unambiguously resolved by lookup in a general or medical domain lexicon. We also do not consider the special case where two period abbreviations are glued together thus forming one token, e.g. “St.p.” = “St. p.” = “Status post” (Latin for “history of”). It furthermore ignores the resolution of ordinal number expressions, for which in German the use of the period character is mandatory, such as “2.” for “2nd”. Consequently, the focus of our investigation is only on non-lexicalised, supposedly ad-hoc abbreviations.

Our hypothesis is that a high efficiency of abbreviation expansion can be obtained in a fully unsupervised fashion, *i.e.* without the (often considerable) effort of producing manually annotated training data. We thus hypothesize that if a clinical collection is sufficiently large, all knowledge needed is present therein.

Materials and Methods

Cardiology corpus

We tested our approach on a corpus of 30,000 pseudonymised discharge summaries from the cardiology department of the Graz University Hospital, the second largest hospital in Austria. The documents were written by German-speaking physicians. Figure 1 shows a typical sample from our corpus.

```

1) Stark vergrößerter li. Ventrikel mit
dünner Wand und höhergr. reduz. systol.
Fkt.; EF: ca. 25- 30%; diffuse
Kontraktilitätstörung; Schaukelbewegung des
LV bei LSB.
2) Li. Vorhof leicht vergrößert;
3) Re. Vorhof leicht- bis mittelgr.
vergrößert;
4) Re. Ventrikel normal groß, normale
Rechtsventrikelfunktion;
5) Aortenklappe morphol. und funktionell
normal;
6) Mitralklappe morphol. unauff.; mittelgr.
MINS; als Mechanismus der MINS besteht eine
Ringdilatation;
7) Trikuspidalklappe unauff.; leichte
TRINS; systol. Pulmonalisdruk ca. 45 mmHg;

```

Figure 1 – Example snippet with ad-hoc abbreviations, e.g. “systol.” for “systolisch” and “Fkt.” for “Funktion”

We first split the corpus into 90% for training and 10% for testing. Two hundred random substrings (100 characters) centred around a period character (followed by a single space) were extracted from the training corpus as source for building a validation set, in which 147 (73.5%) valid abbreviations were manually expanded by the third author, a physician. The period marks in the remaining substrings were considered out of scope for this work (see Introduction for our scope definition). Successive experiments were performed with this set. For the final evaluation, a set of 301 text snippets from the test set was used, of which 200 (66.4%) were considered valid abbreviations. The surrounding context, together with the original text corpus, allowed for unambiguous manual expansions in all cases. We report results from both sets.

N-gram lookup lists

Two frequency tables were created out of the tokenised training corpus, viz. a unigram list U and a bigram list B . The tokenization process ignored period characters, so that they are included in the tokens. Also, as capitalization is an important feature in German and properly used in our corpus, the n-grams were *not* normalised. The lists were arranged in decreasing order of frequency. Corpus frequencies were calculated for all 155,801 token types and all 803,243 bigram types.

Abbreviation and expansion assumptions

An expansion E is a valid expansion of an abbreviation A (with its abbreviation mark, i.e. the final “.” character, stripped) if:

- E does not end with an abbreviation mark (“.”);
- E is at least one character longer than A ;
- E has only alphabetic characters.

Additionally, we define the relative gain $G(A,E)$ as the ratio of the length difference to the abbreviation length, as seen in Equation 1. The intuition behind this restriction comes from the

observation that longer words are rarely abbreviated by overly short strings, as this may lead to ambiguity¹. Nonetheless, they are seldom abbreviated by very long forms either, because the “economy” of using an abbreviation mark would be minimal.

$$G(A,E) = \frac{\text{Length}(E) - \text{Length}(A)}{\text{Length}(A)} \quad (1)$$

Thus, we add extra assumptions regarding relative gains:

- $G(A,E)$ is greater than 0.01;
- $G(A,E)$ is lower than 6.

Finally, two matching types are distinguished, one of which must be true:

- Strict matching: A is a left-sided substring of E ;
- Relaxed matching: all characters of A are contained in E in the same order. The first character of the abbreviation equals the first character of the full form.

While the strict matching accounts for the general case (e.g. “maximal” abbreviated as “max.”), its relaxed version allows the correct matching of “Tbl.” to its expanded form “Tablette”.

Resolution strategy

Our resolution strategy is based on the co-occurrence of adjacent tokens where one of the tokens appears both in shortened and expanded forms in the corpus. As an example, “A. subclavia” can be correctly resolved to “Arteria subclavia”, as the latter is the most common expanded form that matches the abbreviation prefix (see Figure 2). To this end, the bigram frequency list is used. Only when the right or left context does not provide any valid expansions, the unigram frequency list is looked up.

#	Word pair
35	A. subclavia
23	Vena subclavia
12	Arteria subclavia
5	Art. subclavia
4	A subclavia

Figure 2 – Bigram lookup strategy

Considering an abbreviation A , its left context token L and its right context token R , we propose the following algorithm to guess its expanded form E . Every lookup is performed in sequence, hereafter denominated combined approach, until a valid expansion is found.

- Lookup in B for the first bigram LE or ER where E is a valid expansion of A with strict matching;
- Lookup in B for the first bigram LE or ER where E is a valid expansion of A with relaxed matching;
- Lookup in U for the first unigram E where E is a valid expansion of A with strict matching;
- Lookup in U for the first unigram E where E is a valid expansion of A with relaxed matching.

¹ For example, *Pathology* would preferably be abbreviated as *Path.*, as *Pat.* would be ambiguous with *Patient*.

In the special case where any of the contexts L or R is also an abbreviation itself (e.g. “St. p.”), hereafter denominated pairwise abbreviations, two extra initial lookups are performed:

- Lookup in B for the first bigram $E_L E_A$ or $E_A E_R$ where E_i is a valid expansion of i with strict matching;
- Lookup in B for the first bigram $E_L E_A$ or $E_A E_R$ where E_i is a valid expansion of i with relaxed matching.

Similarity index

To overcome mismatches due to adjective inflection endings, which is typical for German, we consider guessed expansions E with similarity $S(A, E)$ to the abbreviation A greater than 70% as a correct match, as seen on Equation 2. The threshold was empirically found in a conservative way.

$$S(A, E) = 1 - \frac{\text{Levenshtein}(A, E)}{\max(\text{Length}(A), \text{Length}(E))} \quad (2)$$

Evaluation

We report precision, recall and F_1 score of our algorithm. Precision is defined as the ratio of correctly expanded tokens to all resolved abbreviations. Recall is defined as the ratio of correctly expanded tokens to all abbreviations [7]. To overcome high precision rates expected when the method gives only expansions with high confidence, F_1 score is defined as the harmonic mean between precision and recall, with equal weights.

Additionally, we define unigram lookup with relaxed matching as a baseline strategy, with which other strategies are compared. We perform a Chi-square test (Fisher’s exact test) to verify if the differences are statistically significant and not by chance, with $\alpha = 0.05$.

Source code

The source code was made available under Version 2 of the Apache License at <https://github.com/michelole/abbres>.

Results

We report results for every lookup considered separately and in the combined approach in Table 1. Bigram strategies, as well as pairwise with strict matching, showed the highest precision rates, but low to moderate recall. When combined with the lower precision unigram approach, we boosted recall to 0.91, thereby achieving an F_1 score of 0.91 in the combined approach (measured in the test set). The combined approach is statistically significant better ($p < 0.001$) than the baseline approach (unigram with relaxed matching). Bigram with relaxed matching is statistically significant better ($p < 0.05$) than the baseline. Any pairwise approach is statistically significant worse ($p < 0.001$) than the baseline approach. Other approaches showed no statistically significant differences to the baseline.

Investigation of errors shows two main patterns: (i) abbreviations preceded or followed by a number (e.g. “52 jähr. Patientin”, which should expand to “52 jährige Patientin”, German for “52 years old patient”); and (ii) abbreviation pairs (e.g. “Z. n.”, which should expand to “Zustand nach”, German for “state after”, i.e. “clinical history of”). We relate the former problem to the lack of any value normalisation (e.g. the conversion of “52” to “00”), which distributes frequencies in the bigram list over every number possibility. Even though we specifically addressed the latter problem in our pairwise resolution strategy, some abbreviation pairs are still ambiguous (“Z. n.” could also

be expanded to “Zeit normal”, German for “normal time”) and a larger context window might be needed.

Discussion

Related work

Several works have been published on the problem of clinical abbreviation detection, expansion, and disambiguation, with different approaches, languages and types of data.

Pakhomov [8] used a semi-supervised Maximum Entropy model for disambiguation of different short forms in a given context. The system was evaluated exploiting a dataset of about 10,000 rheumatology notes from the Mayo Clinic. Extended context information as well as different models trained on disambiguating one specific acronym yielded about the same accuracy of 0.89 for acronym and abbreviation normalization compared to using one model for the normalization task. The hypothesis that similar context information of abbreviations and acronyms compared to their resolved form supports correct normalization in context could be exploited and confirmed in this work.

Joshi et al. [9] compared three different supervised machine learning approaches for acronym expansion: Naïve Bayes Classifier, Decision Trees and Support Vector Machines. All three models achieved an accuracy of over 0.90. The feature set consisted of part-of-speech tags, unigrams, bigrams and a combination of all of them. A flexible window chosen to catch a certain level of occurrence information of lexical features had a significant impact on the overall evaluation efficiency.

Suominen et al. [10] provided an overview of the ShARe/CLEF eHealth Evaluation Lab 2013, in which Task 2 was focused on normalization of abbreviations and acronyms to UMLS concept unique identifiers (CUIs). The challenge used data from the US intensive care publicly available in the MIMIC-II dataset, originally annotated to build the ShARe corpus. The corpus was further enhanced with annotations regarding abbreviations and acronyms text spans and mappings to UMLS codes. The best team obtained an accuracy of 0.72.

Later, Mowery et al. [11] compared the efficiency of participating systems to a majority baseline and with variable majority sense distribution. They observed that a majority approach performed second best (accuracy of 0.69), given that an estimate of around 81% of short forms have no ambiguity (one unique sense) or low ambiguity (two or more senses, with one incidence over 80%). The only winning system showed a slight improvement (accuracy of 0.72) with a hybrid technique that exploits the same differences in abbreviation frequency and ambiguity.

Siklósi et al. [12] addressed the abbreviation resolution problem in Hungarian ophthalmological notes, a language with less linguistic resources available. They specifically dealt with long series of abbreviations commonly found in their documents. The impact of an external lexicon (built out of 3,329 ICD descriptors), a handmade lexicon (with a size of 44 entries) and the corpus itself in the abbreviation expansion phase was evaluated. Additionally, the influence of the context window (from zero to three tokens) and corpus size (with a total of 2,008 documents) in the final expansion efficiency was measured. The expansion strategy leveraged regular expressions built out of the abbreviations and matched (i) again the corpus and (ii) against the lexicons. In the disambiguation phase, their system used a weighted ranking score based on features such as the size of the longest and shortest span covered. We calculated an F_1 score of 0.85 from the reported precision (0.93) and recall

Table 1 – Correct counts (C), precision (P), recall (R) and F_1 score with different strategies and in a combined approach (* denotes $p < 0.05$; ** denotes $p < 0.01$; and *** denotes $p < 0.001$)

Strategy	Training (n = 147)				Test (n = 200)			
	C	P	R	F_1	C	P	R	F_1
<i>Unigram</i>								
Relaxed	91	0.62	0.62	0.62	146	0.73	0.73	0.73
Strict	105	0.76	0.71	0.74	158	0.81	0.79	0.80
<i>Bigram</i>								
Relaxed	119***	0.91	0.81	0.86	165*	0.92	0.83	0.87
Strict	108*	0.94	0.73	0.82	157	0.95	0.79	0.86
<i>Pairwise</i>								
Relaxed ²	24***	0.63	0.16	0.26	49***	0.80	0.25	0.38
Strict	29***	0.91	0.20	0.32	47***	0.90	0.24	0.37
<i>Combined</i>								
	136***	0.93	0.93	0.93	182***	0.91	0.91	0.91

(0.78) metrics. Moreover, their experiments showed that “considering the tokens without any context always performed worst” and “a context larger than one token [...] has a positive effect only if the manually created lexicon is not used”.

Wu et al. [13] applied three different neural word embedding models, viz. SBE (surrounding based embedding feature), LR_SBE and MAX_SBE as additional features for a support vector machine to disambiguate abbreviations in context. The investigation used annotated abbreviation datasets from Vanderbilt University Hospital’s admission notes and narratives from the University of Minnesota-affiliated Fairview Health Services. The MIMIC-II corpus was used to initialise the word embeddings using the algorithm proposed by Collobert et al. [14]. About 42,000 sentences with resolved abbreviations were used for evaluation, achieving a maximum accuracy of 0.96.

Wu et al. [15] developed an open-source framework for clinical abbreviation recognition and disambiguation and evaluated it with a corpus from the Vanderbilt University Medical Center (VUMC) and in the ShARE/CLEF 2013 challenge corpus. They applied semi-supervised clustering methods for sense expansion and profile-based word sense disambiguation. While the former depends on manual review of around 20 sense clusters for each abbreviation to create a sense inventory, the latter builds upon feature vectors representing different senses in a vector space model. Their system achieved an F_1 score of 0.76 in the VUMC corpus, and a 0.29 F_1 score in the ShARE/CLEF dataset.

Our method differs from most of the works being totally unsupervised. Compared to Siklósi et al. [12], the main difference lies in the fact that our disambiguation strategy is also fully data driven and relies only on the frequencies of words and bigrams in a closely related corpus. Although we agree that a weighted ranking of features (e.g. combining the n-gram frequency to its type and relative gain) could improve the disambiguation process, it would need additional annotations to optimize the weight coefficients, thereby transforming our strategy into a supervised approach. Additionally, we did not observe the same incidence of abbreviation series in our corpus, which could explain our better results. Finally, we hypothesize that our much larger corpus (30,000 versus 2,008 documents) might have overcome the need of any lexicon.

Limitations

Our data-driven strategy might show suboptimal results in languages and subdomains with fewer data. It is also sensitive to spelling errors (e.g. “bds.”, German abbreviation for “both sides”, is at least once incorrectly written as “bdsI” in the corpus

— note that the “I” character is found near the “.” character in most keyboards). Although a minimum frequency could be enforced, preliminary experiments showed no improvement over training data.

Furthermore, we propose some basic writing guidelines that could further improve the automated processing of clinical notes. Considering the abbreviation expansion and disambiguation problem alone, guidelines should stress the importance of (a) correct capitalization (e.g. “Patient” instead of “patient”); (b) avoiding typos (e.g. “bdsI” instead of “bds.”); (c) marking abbreviations with a period mark consistently (e.g. “Tbl.” instead of “Tbl”); (d) standardising double abbreviations, preferably with a space character (e.g. “St. p.” instead of “St.p.”); and (e) separating numbers from their units (e.g. “10 mg.” instead of “10mg.”). These simple measures would ease data-driven approaches like ours by increasing the signal-to-noise rate.

Future work

Future work could start exploring the impact of value normalization (i.e. the transformation of all numbers to a standard value) in the results. At least in the training set, this was noted as a common error pattern.

Explorative work also suggests that our algorithm could be useful for the expansion of abbreviations without the period mark, as well as for expanding non-lexicalized and ambiguous acronyms. When uppercased and stripped of their abbreviation mark, abbreviations are similar to acronyms considering the relaxed matching strategy. However, many acronyms are rarely ever expanded, so that the clinical corpus might lack enough full forms. Therefore, other corpora may be necessary, thus complicating the disambiguation task. Our decision to focus on period abbreviations also makes it difficult to compare the results to other works that considered all kinds of short forms.

The proposed strategy should also be evaluated in the broader context of a natural language processing pipeline, in which sentence detection, tokenization, and abbreviation detection are performed together. The special case of glued (without a space mark) pairwise abbreviations (e.g. “St.p.”) might be better addressed with the correct output of a tokenizer. Nonetheless, correct abbreviation detection might be fundamental to distinguish between sentence delimiters and real abbreviations, thereby avoiding false positives.

Apart from evaluation, the application of our method over a full sentence should be considered in the context of a Hidden Markov Model. Hence, the best sentence outcome might be obtained via a dynamic programming algorithm such as the Viterbi algorithm. Larger windows, e.g. trigrams, could be then

² The inclusion of the relaxed pairwise strategy did not improve results in the combined approach and is therefore excluded from it.

exploited to assure maximal context. Suavisiation (e.g. Good-Turing) might be needed for cases unseen in the training set.

Finally, we would like to evaluate our strategy in other subdomains, other languages (e.g. English and Portuguese), and with different corpora sizes, both in publicly available and other restricted corpora. However, our approach would be difficult to run in challenge datasets, like ShARe/CLEF 2013, because it needs a large related corpus as a key resource. For its practical use in clinical language processing this should not be a problem because electronic health record systems do not lack large amounts of text and the choice of domain specific corpora (e.g. cardiology, intensive medicine, nursing, radiology) can be done with attached metadata.

Conclusions

We presented a completely unsupervised approach to the problem of abbreviation expansion. We focused on non-lexicalised and ambiguous abbreviations, commonly created ad-hoc and therefore abundant in clinical narratives. Our strategy is based on bigram and unigram lookup and yielded a 0.91 F_1 score when evaluated with a German cardiology corpus. The result is statistically significant better ($p < 0.001$) than a baseline approach. Our hypothesis that high efficiency rates can be obtained in an unsupervised fashion was therefore not rejected.

Hence, our work provides a successful and reusable method for abbreviation expansion. It improves text comprehension by non-experts and is supposed to improve processing of clinical texts, such as concept mapping and semantic analysis.

Acknowledgements

Our work is funded by the Brazilian National Research Council - CNPq (project number 206892/2014-4). It has been carried out as part of the IICCAB project (Innovative Use of Information for Clinical Care and Biomarker Research) within the K1 COMET Competence Center CBmed, which is funded by the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT); the Austrian Federal Ministry of Science, Research and Economy (BMWFW); the Austrian state of Styria (Department 12, Business and Innovation); the Styrian Business Promotion Agency (SFG); and the Vienna Business Agency. The COMET program is executed by the Austrian Research Promotion Agency (FFG).

References

- [1] H. Xu, P.D. Stetson, and C. Friedman, A study of abbreviations in clinical notes, in: AMIA Annu. Symp. Proc., 2007: p. 821.
- [2] S.M. Meystre, G.K. Savova, K.C. Kipper-Schuler, J.F. Hurdle, and others, Extracting information from textual documents in the electronic health record: a review of recent research, *Yearb Med Inf.* **35** (2008) 128–144.
- [3] J.J. Berman, Pathology abbreviated: a long review of short terms, *Arch. Pathol. Lab. Med.* **128** (2004) 347–352.
- [4] M. Wiesenauer, C. Johnner, and R. Röhrig, Secondary use of clinical data in healthcare providers—an overview on research, regulatory and ethical requirements, *Stud. Heal. Technol. Inf.* **180** (2012) 614–618.
- [5] M. Kreuzthaler, and S. Schulz, Detection of sentence boundaries and abbreviations in clinical narratives, *BMC Med Inf. Decis Mak.* **15 Suppl 2** (2015) S4. doi:10.1186/1472-6947-15-s2-s4.
- [6] M. Kreuzthaler, M. Oleynik, A. Avian, and S. Schulz, Unsupervised Abbreviation Detection in Clinical Narratives, in: Proceedings Clin. Nat. Lang. Process. Work. (ClinicalNLP 2016), Osaka, Japan, 2016: pp. 91–98.
- [7] T. Cohn, Performance metrics for word sense disambiguation, in: Proc. Australas. Lang. Technol. Work., 2003: pp. 86–93.
- [8] S. Pakhomov, Semi-supervised maximum entropy based approach to acronym and abbreviation normalization in medical texts, *40th Annu. Meet. Assoc. Comput. Linguist.* (2002) 160–167. doi:10.3115/1073083.1073111.
- [9] M. Joshi, S. Pakhomov, T. Pedersen, and C.G. Chute, A comparative study of supervised learning as applied to acronym expansion in clinical reports., *AMIA Annu. Symp. Proc.* (2006) 399–403.
- [10] H. Suominen, S. Salanterä, S. Velupillai, W.W. Chapman, G. Savova, N. Elhadad, S. Pradhan, B.R. South, D.L. Mowery, G.J.F. Jones, and others, Overview of the ShARe/CLEF eHealth Evaluation Lab 2013, in: Int. Conf. Cross-Language Eval. Forum Eur. Lang., 2013: pp. 212–231.
- [11] D.L. Mowery, B.R. South, L. Christensen, J. Leng, L.-M. Peltonen, S. Salanterä, H. Suominen, D. Martinez, S. Velupillai, N. Elhadad, G. Savova, S. Pradhan, and W.W. Chapman, Normalizing acronyms and abbreviations to aid patient understanding of clinical texts: ShARe/CLEF eHealth Challenge 2013, Task 2., *J. Biomed. Semantics.* **7** (2016) 43. doi:10.1186/s13326-016-0084-y.
- [12] B. Siklósi, A. Novák, G. Prőszéky, B. Siklósi, A. Novák, and G. Proszéky, Resolving abbreviations in clinical texts without pre-existing structured resources, in: Fourth Work. Build. Eval. Resour. Heal. Biomed. Text Process. Lr., 2014.
- [13] Y. Wu, J. Xu, Y. Zhang, and H. Xu, Clinical Abbreviation Disambiguation Using Neural Word Embeddings, *Proc. BioNLP 15.* (2015) 171–176.
- [14] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, Natural language processing (almost) from scratch, *J. Mach. Learn. Res.* **12** (2011) 2493–2537.
- [15] Y. Wu, J.C. Denny, S. Trent Rosenbloom, R.A. Miller, D.A. Giuse, L. Wang, C. Blanquicett, E. Soysal, J. Xu, and H. Xu, A long journey to short abbreviations: developing an open-source framework for clinical abbreviation recognition and disambiguation (CARD), *J. Am. Med. Inform. Assoc.* **24** (2017) e79–e86. doi:10.1093/jamia/ocw109.

Address for correspondence

Michel Oleynik
Institute for Medical Informatics, Statistics and Documentation
Medical University of Graz, Austria
<http://orcid.org/0000-0001-9099-6259>
michel.oleynik@stud.medunigraz.at

Contrast Mining for Pattern Discovery and Descriptive Analytics to Tailor Sub-Groups of Patients Using Big Data Solutions

Michael A. Phinney^a, Yan Zhuang^b, Sean Lander^b, Lincoln Sheets^{b,c}, Jerry C. Parker^c, Chi-Ren Shyu^{a,b,c}

^a Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, USA,

^b Informatics Institute, University of Missouri, Columbia, Missouri, USA,

^c School of Medicine, University of Missouri, Columbia, Missouri, USA

Abstract

The shift to electronic health records has created a plethora of information ready to be examined and acted upon by those in the medical and computational fields. While this allows for novel research on a scale unthinkable in the past, all discoveries still rely on some initial insight leading to a hypothesis. As the size and variety of data grows so do the number of potential findings, making it necessary to optimize hypothesis generation to increase the rate and importance of discoveries produced from the data. By using distributed Association Rule Mining and Contrast Mining in a big data ecosystem, it is possible to discover discrepancies within large, complex populations which are inaccessible using traditional methods. These discrepancies, when used as hypotheses, can help improve patient care through decision support, population health analytics, and other areas of healthcare.

Keywords:

Data Mining; Electronic Health Records; Population Health.

Introduction

Within the past two decades health records have moved from the paper realm to the digital [1]. As the amount of digitized health data has grown, data mining has been used to make sense of it all. Knowledge that took a lifetime of observation to gain is now obtainable in an instant, given enough data, helping hospitals and researchers to improve their quality of care [2]. For example, the areas of decision support and intervention have both benefited greatly from the application of data mining [3]. Knowledge can now be learned directly from electronic health records using algorithms such as decision trees, association rule mining (ARM), or other pattern recognition systems then acted on by care givers, doctors, and researchers [4]. While this has been a boon to healthcare both at the individual and institutional levels, it has caused a shift in medical research from hypothesis driven to data driven, often removing the hypothesis of “why” from the equation [5]. This “why” is important, however, as treating the symptoms does not always treat the root issue, and so using this knowledge to form hypotheses becomes an important task.

As the volume and variety of healthcare data grows, so does the computational power required to perform analysis. Though a small single-provider practice may be able to run data mining against their data on a single machine, large hospitals or government databases hold too much data for a single machine. ARM, for example, increases exponentially with the number of attributes available. Given n attributes, there are $2^n - 1$ unique combinations of those attributes [6]. Distributed computing techniques allow us to address this exponential scaling factor.

Cluster computing enables distributed analysis of data and storage of large datasets by utilizing an array of machines. Performance can be improved by adding more machines [7]. Through distributed computing in a big data ecosystem it is possible to utilize ARM on datasets on which it would not be possible otherwise. There are several reasons to apply ARM, a special case of Pattern Mining [8], in healthcare applications. Pattern Mining algorithms work by finding groups of items/events/attributes that appear together at higher than expected frequencies. This differs from more traditional statistical and machine learning techniques that may not scale well or have limited explanatory power and intuitively understandable action plans in medicine. While traditional techniques often evaluate a large number of conceivable combinations of attributes, Pattern Mining continuously filters its patterns so that only those which are significant (user defined) are ever evaluated [6]. Using these approaches on health data can still generate millions of rules, many of which can be too general or specific for the research at hand.

While studying each rule on its own may be preferable, it is untenable due to the exponential number of co-occurrence of comorbidities and demographic information. It is necessary to define a way to filter these hypotheses based on their importance and impact. To solve this issue, human-directed hypothesis generation can once again come into play through a process known as Contrast Set Mining (CSM). CSM is the process of using Pattern Mining across a partitioned population in order to find differences in their pattern distributions [9]. Used in its most basic sense, it can be a tool for classification and prediction. In the last decade, CSM has been applied to many data-rich areas, such as genome wide association studies, or disparities in preventative healthcare [10]. Although the ability to use CSM for classification is powerful, patterns alone do not explain why those differences exist or why and how they are important. This, combined with the number of patterns which must be compared, creates four challenging research problems: removing redundant or insignificant patterns, determining the comparative importance of the patterns, doing both at a large scale, and determining why these patterns exist.

One of the primary advantages of using ARM and CSM is the explainability of results. The patterns detected are clearly defined. This is one of the main limitations of many state-of-the-art machine learning algorithms such as support vector machines (SVM) and deep learning on Artificial Neural Networks (ANN); the explanation for classification is convoluted by highly complex models. In a clinical setting, it is important to have transparency in any decision making process. Although decision trees also provide understandable results, they form decisions by considering a single feature a time in a greedy fashion; whereas CSM may consider arbitrarily large

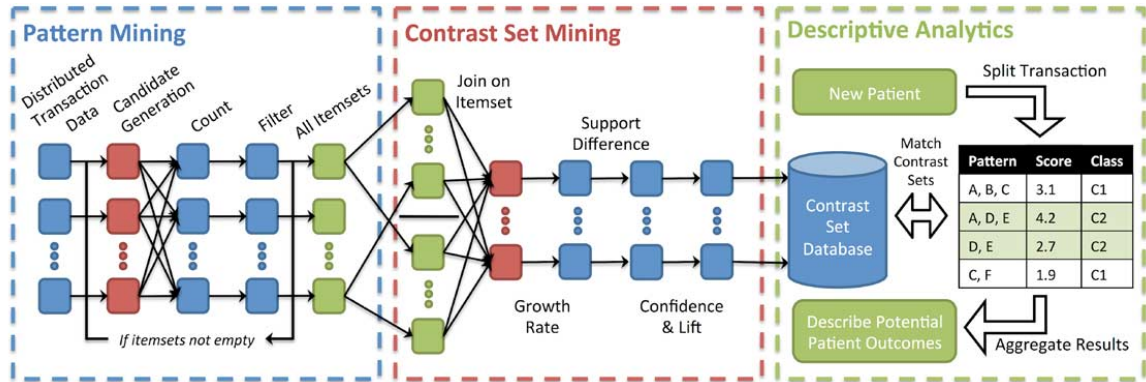


Figure 1 - Pattern mining, contrast mining, and descriptive analytics using a big data approach. Transactions (records) and item sets (combinations of demographic and healthcare information) are distributed across nodes with calculations done in parallel unless aggregation is required

combinations of features at once.

Methods

Pattern mining, at its most basic, can be thought of as applied counting. It is used to find a collection of attributes that occur together above a user-defined frequency. The data, D , can be described as a list of transactions, T . In this research, a patient is considered as a transaction. Each transaction contains a set of items, i , each of which exists in the set of all possible items, I . Patients' demographic information, diagnosis-related groups, and risk levels are potential items for a transaction. Any arbitrary collections of items, $\{i\}$, is considered an itemset. Itemsets with a support above the user-defined minimum are considered *frequent*, with support calculated as $sup(x) = \frac{count(x)}{|D|}$. While the Apriori approach [6] works well for finding common patterns, it quickly exhausts the resources of a normal machine when rare patterns are sought. Common patterns may be found even with a high support; however, rare patterns require a very low support threshold, allowing for near exponential intermediate pattern generation and requiring either numerous reads or very large amounts of memory.

In addition to the volume challenge, the variety of data generates a large amount of intermediate data. During the exploratory analysis, the number of potential itemsets for our data collection reached hundreds of millions, requiring terabytes of memory to index. In order to generate as many patterns as possible, we utilized a Big Data environment to handle large datasets and distributed approaches, which are necessary to tackle scaling challenges. To handle intermediate data of this magnitude, we developed a suite of distributed Apriori tools built on Apache Spark as well as the Apache Hadoop Distributed File System (HDFS) to store the initial, intermediate, and final data. We utilized several low-cost commodity machines equivalent to a system with 18 cores and 144 GB of RAM. For this study, the computing environment consisted of a cluster of 9 Intel NUC machines, each equipped with 16GB of RAM, an Intel i3 processor, and 1.5TB hard drive. We were able to perform the computations by distributing the data across all nodes and performing each step of pattern mining on the data. Most computations could be done in parallel, such as candidate generation and filtering for minimum support, while counting was done by aggregating across the cluster (Figure 1).

In association rule mining, the next process would be to generate rules based on confidence. Rules are generally of the form $\{a, b\} \rightarrow \{c\}$, with the confidence calculated as

$$confidence(\{a, b\} \rightarrow \{c\}) = \frac{sup(\{a, b, c\})}{sup(\{a, b\})}$$

Rule mining can be over-generative when used for classification between groups, with most rules either lacking the class label or with the class in the antecedent. In cases such as these Contrast Set Mining (CSM) can be used instead, a special case of ARM in which the class is always the consequent [9]. CSM is normally applied during ARM in order to only produce patterns which strongly indicate a class. Patterns that have a measurable difference in support are ranked to find those of highest importance. Many different techniques exist for filtering and prioritizing contrast sets, each with its own strengths and weaknesses. All rely on some measure of difference in support. This lends itself well to a distributed approach, as the millions of contrast sets can be grouped, measured, and filtered, allowing even large datasets to be instantly processed and analyzed in parallel for quick data discovery as depicted in Figure 1.

A key goal was reporting mining results that were statistical significant or worthy of conducting further research for vulnerable populations. Patterns rejecting the null hypothesis $H_0: support(X, Population_1) = support(X, Population_2)$ using the Z-Test with high contrasts were selected. Some patterns included minority populations with sample sizes too small to produce an acceptable p values (<0.05) were also selected, to retain visibility into these vulnerable populations.

In our implementation, we designed the system to flexibly apply various filters for different goals after removing non-significant patterns. By setting a target class we could focus on those patterns that had a stronger support in one class than the other. This filtering mechanism allowed us to focus on patterns that were more common in patients that experienced drastic declines in health. It is important to choose the correct measures of significance and importance based on the types of patterns sought. For this we used three significance measures: *growth rate*, *largeness* (support difference), and *confidence*. Though these measures could be calculated in parallel, sorting could not. This final step was done on a single node as shown in Figure 1.

Growth rate is represented as the ratio between supports. This measure is preferred for rare itemsets. The growth rate represents the idea that as more items are added to a pattern, the supports get smaller, and thus differences which may have been imperceptible with smaller, common itemsets become more pronounced as they grow [14].

$$growthrate(X) = \max_{i \neq j} \frac{sup(X, Class_i)}{sup(X, Class_j)}$$

Largeness is represented as the difference between two supports as shown in the following equation. It is most useful when dealing with large supports, a difference between 80% and 60% is greater than the difference between 1% and 5%.

$$largeness(X) = |sup(X, Class_i) - sup(X, Class_j)|$$

Any pattern with a support difference greater than a user-defined value δ is known as large. As mentioned previously, patterns with statistically different ($p < 0.05$) or existent supports are known as significant. Those patterns which are both significant and large are known as deviant [15]. As largeness works with high valued supports, the patterns generated may have a relatively small growth rate.

Confidence is the conditional probability of a $Class_i$ given a certain pattern. That is to say, how strong is pattern X at predicting $Class_i$. The following equation defines the confidence metric, *freq* counts the occurrences of an itemset, and the denominator sums to the total frequency count of pattern X in both classes.

$$conf(X \rightarrow Class_i) = \frac{freq(X, Class_i)}{freq(X, Class_i) + freq(X, Class_j)}$$

Once these measures had been calculated, a post-processing step was run so that only closed patterns remained, a pattern being closed if there is no super-pattern with the same support [16]. This works because it is impossible to increase support by adding items, as the maximum support of the new pattern is the minimum support of all subsets.

$$closed(X): \forall Z \supset X \quad sup(X) > sup(Z)$$

Finally, we obtained three different top-ranked lists, sorted by growth rate for rare contrast sets, or support difference for common, and confidence for both. These three ranked lists of contrast sets provided a broad range of hypotheses to study and act on for healthcare improvement. In order to ensure due diligence, all results generated using these methods were systematically validated against the raw data, ensuring all rules were accurately portrayed.

Experiment Design

The population this data comes from is the LIGHT² (Leveraging Information Technology to Guide Hi-Tech and Hi-Touch Care) project, with goals of improving patient health through risk detection, utilization prediction, prevention, and intervention. Patients in the population were primary care patients in the University of Missouri Health System as well as enrolled in Medicare or Medicaid. Patients were enrolled between February and July of 2013, with 9,581 patients still enrolled by the time the first risk tier evaluation was given on October 1, 2013. All data on the patients' diagnoses, outpatient visits, and hospital visits were based on the University of Missouri Health System electronic medical record, as maintained by clinicians between 2012 and 2014.

For this case study, the attributes used were age (under 65, over 65), sex, marital status, ethnicity, race, language, and 42 diagnosis-related group (DRG) codes applied to the patient over the previous year. This gave a minimum possible itemset size of seven for every patient, as each patient had at least one chronic condition. This resulted in 2^{48} (>280 Trillion) possible attribute combinations for exploratory analyses.

Table 1 - Tier definitions for the LIGHT² project, based on hospital utilization and number of chronic conditions

Tier	Definition (based on past 12 months)
1 Healthy	Chronic conditions defined by the Centers for Medicare & Medicaid Services (CMS) = 0
2 Chronic Stable	Chronic conditions ≥ 1 AND (hospitalizations = 0 AND outpatient visits < 5)
3 Chronic Unstable	Chronic conditions ≥ 1 AND (hospitalizations = 1 OR outpatient visits from 5 to 12)
4 Complex Care	Chronic conditions ≥ 1 AND (hospitalizations > 1 OR outpatient visits > 12)

One essential method of the LIGHT² project was its risk tier schema, separating patients into one of four tiers based on healthcare utilization (Table 1). The lowest, or "Healthy," tier is Tier 1, patients who had not been diagnosed with any of the 27 CMS-defined chronic conditions. For patients with any chronic conditions, Tiers 2, 3, and 4 were based on the number of outpatient clinic visits and hospital visits during the prior 12 months. Utilization was measured separately for (a) outpatient visits, which may be part of normal care management, and (b) hospital-based visits, which usually entail emergency visits, observational stays, or hospital admission. Patient tiers were recalculated every two weeks.

Table 2 - Population statistics for patients stable in Tier 2, and those who moved from Tier 2 to Tier 4. Attributes are a combination of 6 required demographics and DRG codes

Population	Gap Days	Size	Min Attrs	Avg Attrs	Med Attrs	Max Attrs
Stable	N/A	2854	7	10.3	10	23
High Risk	30	170	7	9.5	9	17

In order to generate potentially actionable data, we chose to focus on those patients who started in Tier 2, "Stable" patients suffering from chronic disease who are otherwise healthy. From here, we compared patients who never left Tier 2 with those who left Tier 2 and entered Tier 4 within a 30 day period, indicating a catastrophic worsening of health (Figure 2). A 30-day window was chosen because it indicated multiple hospitalizations or a large number of outpatient visits in an abnormally short period of time. A two-week buffer was added between the end of data collection and the tier movement due to tiers recalculations happening every two weeks. Observations from this period included attributes responsible for tier movement, and thus were not actionable.

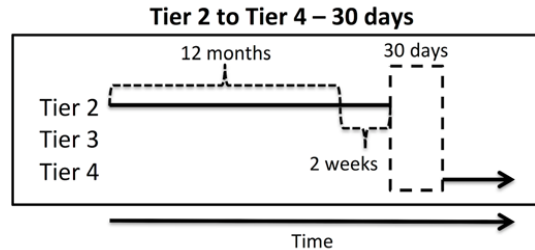


Figure 2 - Population creation criteria used for case study. Those patients who started in Tier 2 and moved to Tier 4 within a 1-month window were chosen as a contrast to patients who started and remained in Tier 2

Due to the patient privacy regulations of the Health Insurance Portability and Accountability Act (HIPAA), detailed statistics regarding this population were not available. However, general statistics about size and complexity of the different populations were available (Table 2). One thing that was surprising when compiling these statistics was that, overall, the population that

stayed in Tier 2 seems to suffer from a higher number of comorbidities. This can be explained by the data collection process, however: while patients who moved from Tier 2 to Tier 4 only had 12 months of data used for analysis, we include data spanning the lifetime of LIGHT2 when analyzing those who stayed in Tier 2. Those patients had a longer period to accumulate DRG codes from visits. Table 3 provides example DRG and demographic codes.

Table 3 - DRG code mappings. This list contains only those codes that appear in the contrast sets in Results

Code	Condition or Attribute
URIN	Bladder, Kidney, and Urinary Tract
IHD	Ischemic Heart Disease (Chronic)
COPP	Conditions Originating in Perinatal Period
CKD	Chronic Kidney Disease (Chronic)
BAA	Black or African American
DIAB	Diabetes

Results

The top ten itemsets that were strongest indicators of being in the high risk patient group, moving from Tier 2 to Tier 4 in a 30 day period, are shown in Table 4. The stable group comprised 94.4% of patients, while the at-risk group was 5.6%. After extracting the strongest contrast patterns based on the growth rate, we used logistic regression to test how well these patterns correlated with the tier movement. The test statistic was a distributed chi-squared with degrees of freedom equal to the differences in degrees of freedom between the current and the null model (i.e., the number of predictor variables in the model). Considering a p value less than 0.05 as a significant model, 8 of the 10 highest growth patterns were significantly correlated with tier movement.

The highest growth rate reported in the first three rows of the table means that the support of the three contrast patterns was 9.593 times greater in the at-risk population. In addition, the highest confidence level was also reported for these three patterns. This confidence 36.4% means that given a contrast pattern, the conditional probability that a patient would move from Tier 2 to Tier 4 in 30 days. This may seem low; however, since the baseline probability for transitioning to Tier 4 was 5.6%, it showed that a patient that exhibits this pattern was over 7 times more likely to transition to Tier 4 than an average patient. The largeness measures were all over 2% which means the percentage of each pattern in Tier movement group was 2% greater than in the stable group.

Another interesting aspect of the results was the size of itemsets; the smallest contrast pattern reported consisted of four items. Within the top 10 contrast patterns there were 12 unique items. In addition to the DRG Codes given in Table 3, we have: over65, 65orLess, Male (M), Female (F), Married, Divorced. Of the 12 attributes, 5 were chronic conditions, 2 were sex, 2 were age, and 1 was race. While chronic conditions were

included in every contrast pattern, marital status occurred in 7, sex in 6, age in 5, and race in 2. The highest support among the contrast sets reported was 2.9% and the lowest support was 2.4%. This suggests the high-risk and stable populations were composed of many subpopulations. Identified contrasts were defined by these small populations.

All reported contrast patterns were indicative of increased risk of hospitalization. The lowest confidence reported was 25%. Although this was a low probability, since the base probability of transitioning to Tier 4 is 5%, it suggests a 5-fold minimum increased risk of transitioning to Tier 4. The support difference was relatively equal between all of the contrast patterns reported in Table 4, around 2%. When all support values are low, the support difference was not as useful.

Discussion

Using these results, it is possible to identify patients who are at risk of having their health deteriorate quickly. Contrast sets formed a strong set of interpretable rules, which can be used by population health managers and clinicians when choosing which patients to spend more time with, and they provide a great starting point for future research questions.

One of the primary advantages of using ARM and CSM is that the results are understandable in a clinical setting, where accountability and transparency are paramount. For example, the top frequent patterns in Table 4 show that married males with chronic kidney disease and ischemic heart disease were significantly more prevalent in the high-risk population. A patient fitting this description is highly likely to have increased hospitalization within 30 days compared to others with stable chronic condition. The third row of Table 4 shows that an African American female over 65 years old with stable diabetes and a history of perinatal conditions also had a high risk of imminent worsening health and hospitalization. Each of these attributes is a well-known risk factor for poor health outcomes; however, these methods identify combinations of risk factors, which lead to particularly high risk of imminent hospitalization. Patients that fit these easily understood profiles could be flagged for additional care management by population health managers and preventive care by clinicians.

One of the main limitations of ARM and CSM is the combinatorial nature of patterns. The methods can create a large amount of data. By utilizing a big data environment, we can extend the limits of this sort of data mining analysis.

Conclusions

With these findings, we have shown how contrast mining on a big data scale can be used on complex datasets for both immediate health care improvement and directing future research in clinical settings. Our findings for the LIGHT² dataset will be used as a guide for patient prioritization by

Table 4 - Top 10 frequent patterns with high growth rate for high-risk population (T2-T4: moved from Tier 2 to Tier 4)

Contrast Patterns (CP)	T2-T4 Support	Growth	Largeness	Confidence	p-value
URIN, M, Married, IHD, CKD	0.024	9.593	0.021	0.364	0.002
Married, CKD, M, IHD	0.024	9.593	0.021	0.364	0.004
F, over65, COPP, DIAB, BAA	0.024	9.593	0.021	0.364	0.199
URIN, over65, Married, IHD, CKD	0.024	6.715	0.020	0.286	0.008
IHD, CKD, over65, Married	0.024	6.715	0.020	0.286	0.016
URIN, Divorced, COPP, M, 65orLess	0.024	6.715	0.020	0.286	0.026
over65, F, DIAB, BAA	0.029	6.457	0.025	0.278	0.177
M, IHD, URIN, Married	0.024	6.105	0.020	0.267	0.004
Married, IHD, URIN, CKD	0.029	5.596	0.024	0.250	0.004
Married, IHD, CKD	0.029	5.596	0.024	0.250	0.008

population health managers, as well as helping project members find new directions to pursue. The combinations of attributes that indicate risk should be provided to population health managers. While this is being done, clinical researchers can formulate more detailed explanations as to why certain individuals are prone to experience severe health complications. By identifying at-risk individuals earlier, the overall quality of care patients receive may be improved.

There are many other large, complex, and new healthcare related areas where this method can be applied. This method can be applied in areas of disparity, adherence, genetic health, and impact of comorbidities. The scalability of distributed computing gives researchers the ability to study larger, more complex datasets than in the past, while this approach to Contrast Set Mining allows for quick exploratory analysis of data as a post-processing technique. The combination of these attributes removes the complexity involved in multivariate analysis and allows for rapid discovery with the need for reprocessing, making distributed Contrast Set Mining an important tool for clinical and translational research.

The next phase of this work is to construct a contrast set classifier for improved prediction of at-risk patients. This data set is particularly challenging due to the skewness between stable and high-risk populations and complex patterns within the high-risk group. Our preliminary results with a direct matching on frequent patterns for the high-risk population are promising and showing advantages over generic machine learning algorithms. Our contrast set classifier was able to achieve a balance between sensitivity and specificity (72% and 32%), while decision tree J48 (0%, 0.1%), random tree (5.2%, 4.7%), Hoeffding Tree (0%, 0%), logistic regression (0%, 0%), and Bayesian Net (0%, 0%) struggled for the high-risk populations due to the complex patterns shared by subgroups of high-risk patients; as we have shown in this paper, the predictive power lies in the combination of key attributes. Our ongoing effort to correctly classify the stable group (patients who stayed in Tier 2) focuses on constructing an aggregate metric to assess which population has the strongest indications of membership in the group. This metric would likely be a combination of growth rate, support difference, confidence, or other contrast metrics.

Acknowledgements

The project was supported by the Department of Health and Human Services, Centers for Medicare & Medicaid Services (1C1CMS331001-01-00). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services or any of its agencies. MAP, YZ, and SAL are co-first authors and with CRS are partially supported by the Shumaker Endowment for Biomedical Informatics. MAP was supported by US Department of Education GAANN Fellowship (P200A100053). The HPC environment is supported by the National Science Foundation (CNS-1429294). We would also like to thank Kayson Lyttle, Julie Jaddoo, and Raymond Nguyen for data collection, cleansing, and preliminary analyses for the project.

References

- [1] S.H. Walsh, The clinician's perspective on electronic health records and how they can affect patient care, *BMJ* **328** (2004):1184-1187.
- [2] E.F. Petricoin, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, and E.C. Kohn, Use of proteomic patterns in serum to identify ovarian cancer, *Lancet* **359** (2002):572-577.

- [3] P.R. Harper, A review and comparison of classification algorithms for medical decision making, *Health Policy* **71** (2005): 315-331.
- [4] R. Bellazzi, B. Zupan, Predictive data mining in clinical medicine: current issues and guidelines, *Int J Med Inform* **77** (2008):81-97.
- [5] D.J. Hand, Statistics and data mining: intersecting disciplines, *ACM SIGKDD Explorations Newsletter* **1** (1999):16-19.
- [6] R. Agrawal, R. Srikant, Fast algorithms for mining association rules, in: Proc. 20th int. conf. very large data bases, *In Proc. 20th int. conf. very large data bases, VLDB* **1215** (1994):487-499.
- [7] R. McCreddie, C. Macdonald, I. Ounis, MapReduce indexing strategies: Studying scalability and efficiency, *Inform Process Manag* **48** (2012):873-888.
- [8] M. Silver, T. Sakata, H.C. Su, C. Herman, S.B. Dolins, M.J. O Shea, Case study: how to apply data mining techniques in a healthcare data warehouse, *J Healthc Manag* **15** (2001):155-164.
- [9] P.K. Novak, N. Lavrač, G.I. Webb, Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining, *J Mach Learn Res* **10** (2009):377-403.
- [10] V.L.S. Thompson, S. Lander, S. Xu, C.R. Shyu, Identifying key variables in African American adherence to colorectal cancer screening: the application of data mining, *BMC Public Health* **14** (2014):1173.
- [11] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, 2016.
- [12] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, Data mining with big data, *IEEE T Knowl Data En* **26** (2014):97-107.
- [13] K. Shvachko, H. Kuang, S. Radia, R. Chansler, The hadoop distributed file system, In: *IEEE 26th Symposium on Mass Storage Systems and Technology* (2010):10.
- [14] G. Dong J. Li, Efficient mining of emerging patterns: Discovering trends and differences, In: *Proceedings of the fifth ACM SIGKDD int. conf. on Knowledge discovery and data mining*, ACM (1999):43-52.
- [15] S.D. Bay M.J. Pazzani, Detecting group differences: mining contrast sets, *Data Min Knowl Disc* **5** (2001):213-246.
- [16] R.J. Bayardo Jr, Efficiently mining long patterns from databases, *Sigmod Rec* **27** (1998):85-93.
- [17] Chronic Conditions Data Warehouse [Condition Categories, , 2017]. Available from: <https://www.ccwdata.org/web/guest/condition-categories>.
- [18] L. Sheets, L. Popejoy, M.K. GCNS-BC APRN, G. Petroski, J.C. Parker, Identifying patients at risk of high healthcare utilization, In: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association (2016):1129.

Address for correspondence

Chi-Ren Shyu, PhD. shyuc@missouri.edu
 Director and Shumaker Endowed Professor of Informatics Institute
 University of Missouri, Columbia, MO, 65211, USA

Detecting Adverse Drug Event Signals from a Clinical CaseBase

Weifeng Qin^a, Xian Zeng^a, Zheng Jia^a, Xiang Zheng^a, Huilong Duan^a, Xudong Lu^a, Haomin Li^{b,c}

^a College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

^b The Children's Hospital, Zhejiang University, Hangzhou, China

^c The Institute of Translational Medicine, Zhejiang University, Hangzhou, China

Abstract

With the rapid development of medical information systems in Chinese hospitals over the last two decades, many of these organizations have accumulated astronomical amounts of structured and unstructured clinical data, including patient diagnostic data, treatment data, lab test data, etc. Secondary use of these data for research, such as Real World Evidence (RWE) studies, has the potential to improve medical quality and safety in daily clinical practice. In this study, we describe CaseBase, a Clinical Data Warehouse (CDW) that extracts structured clinical symptoms or findings and related temporal information from narrative clinical documents and integrates medication information from the CPOE system. An Adverse Drug Event (ADE) signals detection platform has also been developed based on CaseBase to analyze and visualize drug-symptom relations in clinical data. A prototype of this platform has been evaluated in a 2,000-bed hospital and some initial results are reported here.

Keywords:

Information Storage and Retrieval; Drug-Related Side Effects and Adverse Reactions

Introduction

Secondary use of Real-World Data (RWD) from clinical information systems has been studied for many years, as it has been proven to be beneficial to both medical research and the quality of clinical practice [1]. Particularly, it could help to strengthen our current understanding of healthcare delivery and patient outcomes, as well as increase the potential for generating new knowledge [2, 3]. Furthermore, RWD could provide evidence for clinicians to guide their practice [4, 5]. Despite the above benefits, due to the distributed and heterogeneous nature of RWD, they are difficult to use directly. As a result, most clinical RWD remains underutilized by hospitals. This is a great waste of clinical information resources.

To fully use clinical RWD, it generally requires a prior integration and structural optimization process in a unified clinical database, also called a Clinical Data Warehouse (CDW). Generally, a CDW is used to support medical education or research, or to satisfy demands for management information and clinical information in the medical field [6].

In addition to using RWD to improve care and organizational processes, there are also potential applications in drug safety, which can vary from controlled trial findings when used widely in practice. Each year, ADEs cause a great threat to human life and health worldwide [7, 8], and place a huge economic burden on the world [9-11]. Most of the ADEs are known and have been labeled with the drug specification, but there are still a small number of potentially unknown ADEs. By identifying

these unknown ADEs, ADE signal detection [12] can be used to help reduce the incidence of potentially serious ADEs.

In this study, we describe CaseBase, a secondary use oriented CDW based on RWD from the Computerized Physician Order Entry (CPOE) system and Electronic Medical Record (EMR) in a Chinese hospital. Additionally, we will describe our application of the statistical frequency based Disproportional Analysis (DPA) [13] method to develop an ADE signal detection platform based on this CDW.

Methods

CaseBase Database Design

Given the sparse and dynamic characteristics of the raw clinical data involved in this study, the Entity-Attribute-Value (EAV) modeling methodology is used for the CaseBase database design [14-16]. In this study, one patient visit is seen as one case in the CaseBase, and one case could contain many data items or facts. For example, data items like a symptom, a laboratory test result, or a drug order all could be seen as facts.

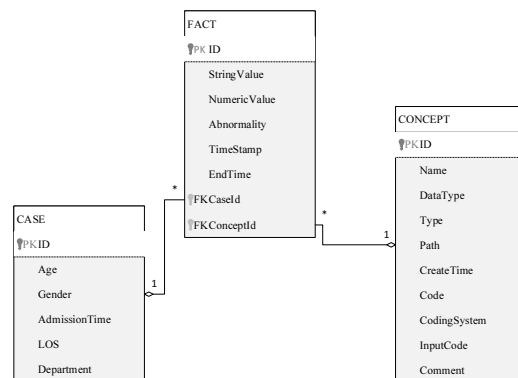


Figure 1-Tables Design of CaseBase

Therefore, the CaseBase database is composed of one FACT table and two other related tables, named the CASE table and CONCEPT table respectively, as shown in Figure 1. These three tables form a star structure, and combined, they describe one fact. Specifically, the FACT table in the center is used to store the detailed core data content of one fact, such as the result of a laboratory test, the name of a drug, the symptom of a patient, etc. The CASE table is designed to store the relevant patient case information about the fact and also can be used to assist in retrieving fact data. To protect the privacy of patients, names, phone numbers, and family addresses are not contained in the CaseBase. Finally, the CONCEPT table is used to store

the detailed description items for a specific fact, such as the name and type of fact. In brief, the CASE and CONCEPT tables are used to support the FACT table.

The ETL Process

The Extract, Transfer, Load (ETL) process will extract the required structured patient fact data from distributed and heterogeneous medical information systems and load them into the CDW. The process includes three steps: Extracting clinical data stored in various information systems, Transforming the extracted data's structure and meaning, and Loading the transformed data into the CaseBase.

In this study, patient medication lists from the CPOE system and relevant progress notes from the EMRs are processed and loaded into CaseBase. All this information can be accessed through an Integration Engine deployed in the hospital. A platform-independent Web Service application called BuildCase was developed to interoperate with the Integration Engine and the CaseBase. For instance, when a patient is discharged from the hospital, the Integration Engine will call the BuildCase service with a parameter containing the patient's VisitId (one patient's unique identifier in one visit to this hospital), then the BuildCase would access the information systems and get the required data.

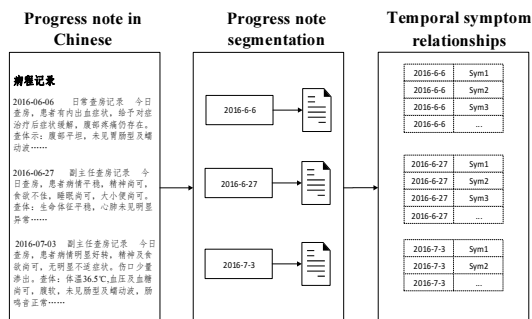


Figure2 - Process of Symptom Extraction

In the second step, the extracted raw clinical data is transformed into the required format, which involves two aspects of transformation, the data structure and the data meaning. For the former, it needs the BuildCase service to restructure the data extracted from the heterogeneous information systems according to the demands of the CaseBase data model. Most of the data coming from the CPOE system require this step. For the latter, it is more complex relative to the data structure transformation, for example, transforming the patient's birth date to age. To get the patient symptom data, a dictionary-based Chinese Named Entity Recognition (NER) technology, a kind of Natural Language Processing (NLP) technology, is applied to transform the free text, unstructured progress notes from the EMR system to the form of symptom terms needed for the timeline. As shown in Figure 2, a patient's progress note was recorded regularly day by day. The progress note was first segmented into multiple parts by time using regular expressions. Then, each part was processed into several temporal symptom relationships using a Chinese NER technology based on a home-grown medical terminology dictionary described in our previous work [17].

In the last step, the transformed data need to be loaded into the CaseBase CDW, and the whole ETL process is finished. In addition, all the ETL process related services are called in asynchronous mode, and thus will not change the data in the

original system and reduce the running speed of the source system.

Mining of the DEC and the DPA method

To mine the potential Drug-Effect Combinations (DECs), some pre-processing and data cleaning steps are carried out ahead of the whole process. Specifically, the facts data unrelated to drugs or ADEs are filtered out first; then the facts that belong to solution type (e.g glucose solution) are filtered out. Next, the long-term doctor's orders are divided into single orders with the day as the unit; lastly, the symptoms that occurred within 12 hours after admission are filtered out, because most of the symptoms that happen in the first 12 hours were not treatment-related.

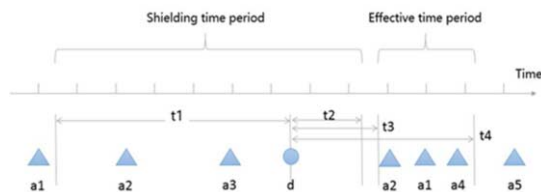


Figure 3 - Judgement Standard of DEC

After the completion of pre-processing, a DEC screening process is applied as illustrated in Figure 3. In the figure, the horizontal axis represents time, the circle point represents that a drug has been used, and the triangle represents a symptom. In order to include only the symptoms occurring after drug use, rather than before drug use, a shielding set A, a candidate set B and a target set Δ were defined, and their detailed definitions are represented by mathematical expressions as follows.

$$A = \{x | x.time \geq d.time - t1 \text{ and } x.time < d.time + t2\}$$

$$B = \{x | x.time \geq d.time + t3 \text{ and } x.time < d.time + t4\}$$

$$\Delta = \{x | x \in B \text{ and } x \notin A\}$$

Thus, according to the above conditions, the DEC could be defined as the two tuples formed by the drug d and the symptom in the Δ set decided by drug d . For example, as shown in Figure 3, the shielding set A contains symptoms a2 and a3, the candidate set B contains symptoms a1, a2 and a4, and as a consequence, the target set Δ contains a1 and a4. So, the DEC's selected are d -a1 and d -a4, two pairs in total. In addition, the four parameters $t1$, $t2$, $t3$, and $t4$ can be adjusted for different study purposes, generating different results.

		Adverse effect	
		Present	Absent
Drug	Present	A	B
	Absent	C	D

Figure 4 - Cross Table Method

On the basis of the DEC's produced above, the DPA method, one of the most commonly used ADE signal detection methods, is applied to this study to mine the potential ADEs. In this method, the frequencies of presence or absence of the drug and adverse effects are recorded based on the cross table method illustrated in Figure 4. As shown in the figure, there are four frequency parameters (A, B, C, and D) defined to count the statistical frequency for each of the DEC's of interest. The letter A represents frequencies for when the drug of interest and the adverse effect of interest are present together. The letter B

represents frequencies for the drug being present and the adverse effect being absent. The letter C represents the frequencies for when drug is absent and the adverse effect is present. The letter D represents the frequencies for when both the drug and the adverse effect are absent. Based on these four statistical frequency parameters, four highly correlated assessment factors can be calculated. They are Relative Reporting (RR), Proportional Reporting Rate Ratio (PRR), Reporting Odds Ratio (ROR), and Information Component (IC) respectively, and their calculating formulas are listed as follows.

1. $RR = \frac{A(A+B+C+D)}{(A+C)(A+B)}$
2. $PRR = \frac{A(C+D)}{C(A+B)}$
3. $ROR = \frac{AD}{CB}$
4. $IC = \log_2 RR$

Results

In this study, a total of 23,898 cases from a 2,000-bed hospital were collected in the CaseBase, from which 8,320 cases were included for ADE signal detection. Particularly, there were 250,865 facts of symptoms and 414,500 facts of orders in these cases. Usually, the finding and recording time of an effect or a symptom of a patient by a physician would be later than the creation time of the drug order; therefore, [24h-72h] was selected as the effective time window, meaning the parameter t3 equals 24 hours and t4 equals 72 hours. Finally, a total of 150,623 DEC of 71,980 different kinds were selected for inclusion.

Further analysis found the frequency of most DEC to be no more than 2, and the kinds of DEC reduced rapidly with increasing DEC frequency. Moreover, the average IC value of

all frequencies showed no obvious difference. According to these analyses, the DEC that have a relatively high frequency and a larger IC value should be more likely to be the ADE signals. Therefore, the DEC in which frequencies are no less than 3 and IC values are above 20 were selected (shown in Table 1). From the table, first, it can be seen that the four factors PRR, ROR, RR, and IC have a high correlation, and second, the 20 DEC could be divided into four categories as listed in the last column.

The “confirmed” category means the side effects have already been discovered in clinical trials or studies, the “indication” category means the DEC is a kind of indication symptom of the drug, the “error” DEC are caused by wrongly extracted symptoms, and the “suspected” DEC are possible new drug-ADE pairs that need to be further analyzed and researched. As can be seen through preliminary analysis, in the 20 DEC, the “suspected” DEC make up about half, and the “indication” and “confirmed” DEC each account for a quarter.

The detailed mining result was visualized using Microsoft Silverlight technology for a project involving our organization’s clinical pharmacist. A sample of the visualization results are shown in Figure 5. The drugs and symptoms are represented as bubbles with different colors. The size of the bubble shows the number of cases in which the drug was used or the number of times a symptom was observed in the data set. The distance between symptom and drug bubbles represents the confidence level measured by the IC assessment factor. Drug and symptom pairs with larger IC values will have a shorter distance. The width of the arrow between drug and symptom bubbles shows the frequency of the drug-symptom pair occurring under this profile. The color of the arrow was used to distinguish whether the relationship is a known side effect of the drug in our knowledgebase, indicated below by a golden arrow.

Table 1 – IC Top 20 DEC

DECs		A	B	C	PRR	ROR	RR	IC	CATEGORY
Etanercept	Injection site reaction	4	45	15	818.3	890.9	646.5	9.34	Confirmed
Carboplatin for Injection	Gastrointestinal discomfort	6	15	90	477.8	668.4	448.1	8.81	Confirmed
Metronidazole Tablets	Dysmenorrhea	3	154	5	574.3	585.5	359.5	8.49	Suspected
Metronidazole and Sodium Chloride Injection	Dysmenorrhea	3	242	5	368.0	372.5	230.4	7.85	Suspected
Cyclophosphamide for Injection	Hemorrhagic cystitis	11	179	27	322.5	342.3	229.4	7.84	Confirmed
Diacerein Capsules	Tenosynovitis	3	47	40	225.7	240.1	210.1	7.71	Indication
Levodopa and Benserazide Hydrochloride Tablets	Extrapyramidal symptoms	3	90	21	231.1	238.8	202.4	7.66	Indication
Vitamin B2 Tablets	Oral ulcer	3	14	129	205.9	249.8	201.3	7.65	Suspected
Metronidazole Tablets	Pelvic inflammation	3	154	13	221.1	225.3	179.8	7.49	Indication
Metronidazole Tablets	Pelvic mass	3	154	13	221.1	225.3	179.8	7.49	Suspected
Metronidazole Tablets	Ovarian cyst	5	152	22	217.8	224.9	177.6	7.47	Suspected
Sterile Water for Injection	Injection site reaction	4	183	15	214.4	219.1	169.5	7.41	Suspected
Chymotrypsin for Injection	Scleroma	15	58	170	182.0	228.7	167.3	7.39	Error
Pirarubicin Hydrochloride for Injection	Gastrointestinal discomfort	3	28	93	156.6	173.3	151.8	7.25	Confirmed
Metronidazole and Sodium Chloride Injection	Ovarian cyst	6	239	21	175.3	179.7	136.6	7.09	Suspected
Aciclovir for Injection	Meningeal irritation sign	3	116	28	135.5	138.9	122.5	6.94	Suspected
Shenkang Injection	Cystopyelonephritis	3	371	7	172.1	173.5	120.8	6.92	Indication
Metronidazole and Sodium Chloride Injection	Pelvic mass	3	242	13	141.6	143.3	115.2	6.85	Suspected
Metronidazole and Sodium Chloride Injection	Pelvic inflammation	3	242	13	141.6	143.3	115.2	6.85	Suspected
Oxybuprocaine Hydrochloride Gel	Rectitis	3	185	19	126.3	128.3	109.2	6.77	Suspected

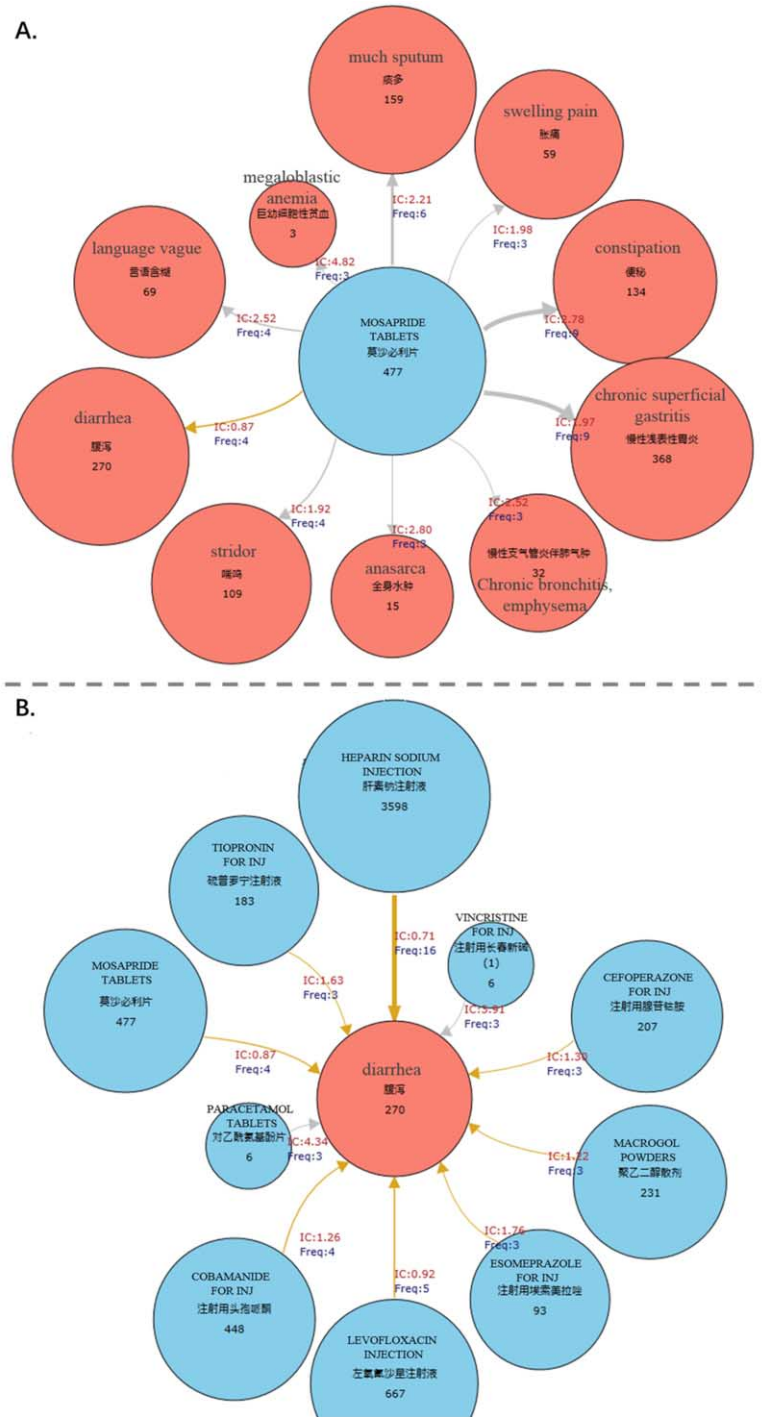


Figure - 5 Sample of the visualization result for drug-symptom relationships: A. Drug-centered view. B. Symptom-centered view.

Furthermore, clicking on any drug or symptom bubble will navigate to a new scenario. For example, clicking the MOSAPRIDE TABLETS in Figure 5B will navigate to the MOSAPRIDE TABLETS centered view as shown in Figure 5A. So users can evaluate both the potential adverse effect of a drug and the potential drugs causing an adverse effect in a user-friendly way.

Discussion

This is the first prototype of an ADE signal detection platform using Chinese EMRs that has been reported. Although only a small dataset was used in this study, it definitely showed its feasibility for RWE studies in clinical information systems in

China. Considering the increasingly serious drug-safety issues in China, we believe this prototype will inspire more investigations in this field.

The most challenging part of this study was to automatically extract symptoms from progress notes, as there are still limited Chinese symptom terminology resources available. To evaluate symptom extraction methods, 100 progress notes were randomly selected from the EMR system and manually annotated by two independent annotators. These two annotated results were combined to create a gold standard for the evaluation. A total 666 symptoms were manually annotated as positive symptoms observed in the 100 progress notes. The NLP methods used in this study extracted 399 symptoms from the test progress notes, of which 388 symptoms met the gold standard. Thus, the recall rate of the NLP methods did not perform very well in the evaluation, although the 97.2% precision rate is high enough to provide a good foundation for further analysis. It is hard for a home-grown medical terminology dictionary to cover all symptoms, thus a comprehensive Chinese symptom terminology dictionary or other machine learning based NER method is needed for future study.

Timestamps of the progress notes were assumed in this study to be the time when symptoms were observed, though we found this is not always the case in certain conditions. To improve the temporal accuracy of extracted symptom facts, a temporal information extraction and relationship extraction study will be conducted in the future.

In the future, if this ADE signal detection platform is deployed in a real-time clinical setting, the facts database will be up-to-date, and analysis results and reports will be generated every day, week, and month to present the potential ADE signals for our organization's clinical pharmacist. This project holds promise to help our clinical pharmacist to detect ADE signals earlier or work as an efficient support platform for our current spontaneous ADE reporting system.

Conclusions

In this study, making use of the clinical data from a 2,000-bed Chinese hospital and the ETL process, a CDW called CaseBase has been constructed to facilitate the secondary use of clinical RWD for the purpose of providing RWE for clinical studies or projects. Additionally, an ADE signal detection platform prototype based on the constructed CDW was developed and some initial results were obtained. These results preliminarily demonstrated how this work is feasible to facilitate the secondary use of clinical RWD.

Acknowledgements

This research was financially supported by the national key research project 2016YFC0901905.

References

- [1] S. L. Mackenzie, M. C. Wyatt, R. Schuff, et al. Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey [J]. *Journal of the American Medical Informatics Association: JAMIA*, 2012, 19(e1): e119-E24.
- [2] M. Pobiruchin, S. Bochum, U. M. Martens, et al. A method for using real world data in breast cancer modeling. *Journal of Biomedical Informatics*, 60 (2016) 385–394.
- [3] T. B. Murdoch, A. S. Detsky. The inevitable application of big data to health care. *JAMA*. 2013; 309:1351–2.
- [4] M. De Boer, E. B. Boeker, M. A. Ramrattan, et al. Adverse drug events in surgical patients: an observational multicenter study [J]. *International Journal of Clinical Pharmacy*, 2013, 35(5): 744–52.
- [5] J. O. Oyindola, J. Campbell, A. A. Kousoulis, Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidelines. *BMC Health Services Research*, (2016) 16:299
- [6] M. De Mul, P. Alons, P. Van Der Velde, et al. Development of a clinical data warehouse from an intensive care clinical information system [J]. *Computer methods and programs in biomedicine*, 2012, 105(1): 22–30.
- [7] L. Lian Duan, M. Khoshneshin, W. N. Street, et al. Adverse Drug Effect Detection [J]. *IEEE Journal of Biomedical and Health Informatics*, 2013, 17(2):305–11.
- [8] J. Lazarou, B. H. Pomeranz, P. Corey, Incidence of Adverse Drug Reactions in Hospitalized Patients: A Meta-analysis of Prospective Studies [J]. *JAMA: The Journal of the American Medical Association*, 1998, 279(15): 1200–5.
- [9] J. A. Johnson, J. L. Bootman. Drug-related morbidity and mortality. A cost-of-illness model [J]. *Archives of internal medicine*, 1995, 155(18): 1949.
- [10] S. Galjtier, H. Bachelet, R. Bordet, et al. The cost of adverse drug reactions [J]. *Expert opinion on pharmacotherapy*, 2003, 4(3): 319.
- [11] 2012 Annual Report of ADR from China Food and Drug Administration [M].
- [12] E. P. Van Puijenbroek, A. Bate, H. G. M. Leufkens, et al. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions [J]. *Pharmacoepidemiology and drug safety*, 2002, 11(1): 3–10.
- [13] T. W. De Vries, J. J. De Langen-Wouterse, E. P. Van Puijenbroek, et al. Reported adverse drug reactions during the use of inhaled steroids in children with asthma in the Netherlands [J]. *European journal of clinical pharmacology*, 2006, 62(5): 343–6.
- [14] W. W. Stead, W. E. Hammond, M. J. Straube. A chartless record – is it adequate? *J Med Syst*, 1983; 7: 103–9.
- [15] C. D. Niedner. The entity-attribute-value data model in radiology informatics. In: *Proceedings of the 10th conference on computer applications in radiology*, Anaheim, CA; 1990.
- [16] P. M. Nadkarni, C. Brandt. Data extraction and ad hoc query of an entity-attribute-value database. *J Am Med Inform Assoc*, 1998; 5: 511–27.
- [17] C. Ge, Y. Zhang, H. Duan, H. Li, Identification of adverse drug events in Chinese clinical narrative text, *International Conference on Information Technology in Medicine and Education*, 2014.

Address for correspondence

Haomin Li, The Children's Hospital, the Institute of Translational Medicine, Zhejiang University, Hangzhou, China,

E-mail: hmli@zju.edu.cn

Extraction of Data from a Hospital Information System to Perform Process Mining

Ricardo Alfredo Quintano Neira^{a,b}, Gert-Jan de Vries^c, Jennifer Caffarel^c, Erin Stretton^c

^aIndustrial Engineering Department, Pontificia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

^bPhilips Research Brazil, Barueri, SP, Brazil

^cPhilips Research - Healthcare, Eindhoven, the Netherlands

Abstract

The aim of this work is to share our experience in relevant data extraction from a hospital information system in preparation for a research study using process mining techniques. The steps performed were: research definition, mapping the normative processes, identification of tables and fields names of the database, and extraction of data. We then offer lessons learned during data extraction phase. Any errors made in the extraction phase will propagate and have implications on subsequent analyses. Thus, it is essential to take the time needed and devote sufficient attention to detail to perform all activities with the goal of ensuring high quality of the extracted data. We hope this work will be informative for other researchers to plan and execute extraction of data for process mining research studies.

Keywords:

Hospital Information Systems; Process Mining; Database Management System

Introduction

Process mining consists of a set of techniques that enable the analysis of business processes using system data. Specialized algorithms treat the data identifying patterns and trends. The use of process mining algorithms allows to discover process models (process discovery), identify deviations in a process (conformance checking), identify bottlenecks and performance indicators (performance checking), and identify how information flows between resources using social networks [1,2].

The event log is the raw material for running process mining algorithms. It contains all events used to construct a journey map and has as main attributes a case ID that represent one instance of the process (in healthcare field it could be the patient or hospitalization identification), the activity performed (e.g. "Perform triage", "Discharge of patient"), and the date and time the activity was performed [1,3].

Several research studies have applied process mining techniques for healthcare and they are present in many medical fields such as cardiology, oncology, diabetes and clinical images [4][5]. For example, Forsberg et al. [6] performed a study to identify the reading chest radiograph process in a Picture Archiving and Communication System (PACS). Rattanavayakorn and Premchaiswadi [7] applied the "working together metric" of social miner techniques to understand the behavior of healthcare professionals when treating patients in a hospital in Bangkok. Mans et al. [8] applied the heuristics miner algorithm to discover and compare the stroke treatment process in four Italian hospitals. Also, they applied performance checking algorithms to

discover the bottlenecks and performance indicators for the pre-hospital process. Huang et al. [9] presented a technique that creates a summary of the structure of clinical pathways. They applied the approach for four different diseases (bronchial lung cancer, colon cancer, gastric cancer, and cerebral infarction) discovering essential medical behaviors with specific execution order. Another study analyzed the control flow, organizational and performance perspectives of a gynecological oncology process in a Dutch hospital to obtain insights in the care flow [10].

We are performing a research study using process mining techniques to identify deviations and bottlenecks in a sepsis treatment process in a Brazilian hospital. We expect to identify actions (changes in the process) that can improve the sepsis treatment process. Our first step was the extraction of data from a Hospital Information System (HIS).

The aim of this work is to present the steps we followed to extract data from the HIS to perform the process mining work. The main purpose of this work is to share our experience regarding the data extraction and all the preparation work associated with this task. We hope that our experience can help other researchers to plan and execute the extraction of data for process mining research studies.

Methods

Below we present all steps we followed to extract the data from the HIS database.

Research definition

First, we defined the research questions we want to answer in our work:

1. Which is the *AS-IS* (current process) sepsis treatment process of the hospital? How does the hospital staff treat septic patients?
2. Which are the deviations in the process? Do professionals perform activities in a different order than defined in the normative sepsis treatment process?
3. Which are the bottlenecks in the process? Are there activities in the process that are taking more time than expected?
4. What is the workload of each professional in the hospital? Could a heavy workload cause delays in the process?
5. Which actions can improve the process? For example, we would like to identify changes in processes that might reduce the time it takes to administer initial antibiotic therapy.

The research questions were crucial to understanding which type of information we should extract from the database. For example, if we want to know the AS-IS process, we need to create an event log with the case identification (in our case it is the hospitalization identification), the activity type ("Registry of patient," "Triage," "Medical evaluation") and the completion date and time. If we want to analyze the workload of healthcare professionals then in the event log we need to add the health professional identification and extract information about all hospitalizations (not only of sepsis) to get a complete overview of health professionals tasks.

Mapping the process

In a first visit to the hospital, we analyzed the process the hospital applies for treating sepsis patients. We studied the hospital documents (sepsis guidelines and sepsis screening form), performed interviews with health professionals (2 physicians, 2 nurses, 2 nurse technicians, 1 quality analyst, 1 receptionist) and performed shadowing. The number of interviewers was four.

With all the information collected, we designed two models using the Business Process Model and Notation (BPMN). One model represents the sepsis treatment in the Emergency Department and the second the sepsis treatment in the Intensive Care Unit. The models represent the way health professionals should work when treating sepsis patients (normative models). Both models were updated and validated by the staff (4 physicians, 3 nurses, 2 nurse technicians, 1 quality analyst, 1 laboratory technician, 2 pharmacists) in a second visit to the hospital. The number of interviewers was five.

Figure 1 presents a simple model of a treatment process based in the Emergency Department model. We created it to help the reader to understand the extraction process. We linked the process activities to the next steps of the extraction.

The process models and all the information collected from the hospital visits were very important to guide us in the selection of the attributes to collect from an immense amount of attributes presented in the database. For example, in the "1. Register Patient" activity we needed to know the time and date this event happened, and who performed the registration.

Identification of tables and fields of the database

When we started this research, we had no knowledge about the HIS database structure. We did not know from which

tables and fields we should extract the data we needed to perform our research.

We asked the HIS development team to give us guidelines of which tables and fields we should collect all the information needed for our research. For this task, we created a spreadsheet containing all attributes that we needed to extract from the database based on the research questions and process models. We sent this spreadsheet together with the process models to the HIS development team and asked them to complete it.

We created two versions of the spreadsheet. The first one presented just one tab with the attributes needed by each step (activity) of the treatment processes. We denominate this tab as "Process Oriented". Since it was difficult and confusing for the development team to fill it, we created a second version of the spreadsheet that contained a new tab, denominated "HIS Modules Oriented", presenting the same attributes needed grouped by functionality of the system. We believe this module oriented view would help them to more easily associate the correct tables and fields of the database, since this is closer to their way of thinking (they do not necessarily know the clinical process that is followed, but they do know the modules used by the users). The main difference between the process and module tabs is the way that the attributes needed are organized. The "Process Oriented" tab presents the attributes grouped by each activity of the process, and the "HIS Modules Oriented" tab presents the same attributes grouped by each module of the HIS. To convert an item from the "Process Oriented" tab to the "HIS Modules Oriented" tab we identified the HIS module used by the clinical user to register the information (attribute) associated with the process activity.

Below we present both tab structures of the spreadsheet.

Process oriented tab

Table 1 presents a small sample of the first tab. It contains the following columns:

- Step name (A): name of the activity from the BPMN model. E.g., "1. Register Patient", "2. Perform Triage" from Figure 1;

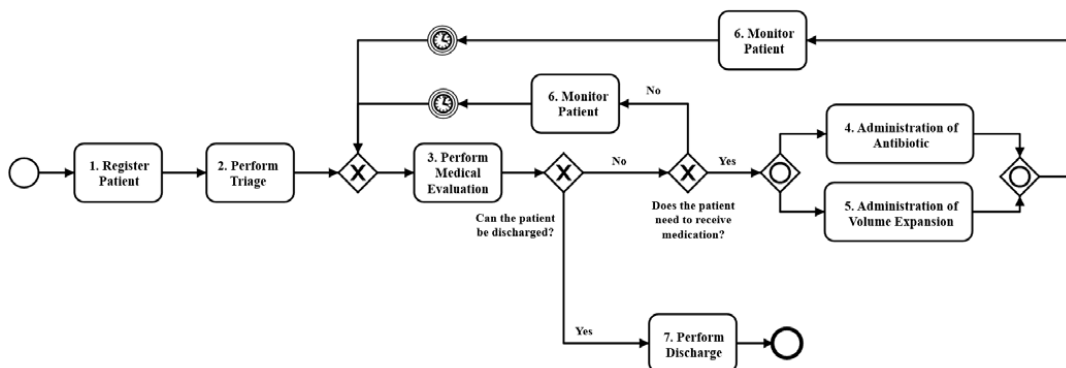


Figure 1 - Simple BPMN model representing a treatment process in the emergency department (Note: this model was created as an elucidative example. It does not represent the exact and complete process as followed by the hospital)

Table 1 - Sample of process oriented tab of the spreadsheet (Note: all content data presented is fictitious. The idea is to present the structure of the table. The table presents a subset of the activities from Figure 1)

Step Name (A)	Attribute (B)	Responsible (C)	Table (D)	Table Field (E)	Execution		Registry	
					Date and Time (F)	Role of User (G)	Date and Time (H)	Role of User (I)
2. Perform Triage	Clinical Notes	ES	TRIAGE	clinical_notes	exec_date	exec_role	end_date	user_role
2. Perform Triage	Temperature	JC	VITAL_SINGS	temperature	start_date	perf_role	reg_date	reg_role
2. Perform Triage	Blood Pressure	JC	VITAL_SINGS	blood_pressure	start_date	perf_role	reg_date	reg_role
3. Perform Medical Evaluation	Clinical Notes	GJV	NOTES	clinical_notes	exec_date	exec_role	end_date	user_role
6. Monitor Patient	Temperature	JC	VITAL_SINGS	temperature	start_date	perf_role	reg_date	reg_role
6. Monitor Patient	Blood Pressure	JC	VITAL_SINGS	blood_pressure	start_date	perf_role	reg_date	reg_role

Table 2 - Sample of HIS module oriented tab of the spreadsheet (Note: all content data presented is fictitious. The idea is to present the structure of the table.)

Module (A)	Attribute (B)	Responsible (C)	Table (D)	Table Field (E)	Execution		Registry	
					Date and Time (F)	Role of User (G)	Date and Time (H)	Role of User (I)
Triage	Clinical Notes	ES	TRIAGE	clinical_notes	exec_date	exec_role	end_date	user_role
Vital Signs	Temperature	JC	VITAL_SINGS	temperature	start_date	perf_role	reg_date	reg_role
Vital Signs	Blood Pressure	JC	VITAL_SINGS	blood_pressure	start_date	perf_role	reg_date	reg_role
Electronic Health Record	Clinical Notes	GJV	NOTES	clinical_notes	exec_date	exec_role	end_date	user_role

- Attribute (B): name of the attribute that we need from the process activity. E.g., from the triage step (2. Perform Triage) we need to extract the “temperature”, “blood pressure”, and “clinical notes”. One step name can have many attributes;
- Responsible (C): contact information of the responsible from the HIS development team who filled the spreadsheet line. This is important in case we had doubts about an attribute and thus we could contact directly the professional;
- Table (D): name of the table from the database where the field is located;
- Table field (E): name of the field from the database which contains the attribute information to be extracted;
- Execution - date and time (F): field from the database that contains the date and time that the action was performed in practice. E.g. “What was the time that the administration of the medication was performed for patient John?”;
- Execution – role of user (G): field from the database that contains the role of the professional who performed the action. E.g. “Nurse”, “Physician”;
- Registry - date and time (H): field from the database that contains the date and time that the attribute was entered in the system. E.g. “What was the time that the administration of the medication for patient John was entered in the HIS?”;
- Registry - Role of user (I): field from the database that contains the role of the professional who entered the attribute in the system. It is important to also retrieve this information as the person documenting the activity may not be the same as the one executing it

(e.g. a doctor may perform an action and ask a nurse to document it in the HIS).

Columns from C to I should be filled by a professional from the HIS development team.

HIS modules oriented tab

Table 2 presents a small sample of the second tab. Columns B to I are the same from Table 1. The column “Step Name” (A) was replaced with “Module”. This new column presents the module name from the HIS that the information requested can be collected. E.g. “Patient registry”, “Electronic Health Record”, “Computerized Physician Order Entry”, “Imaging”, “Transfer of Patient”. Columns from C to I should be filled by a professional from the HIS development team.

Table 1 and Table 2 present the same content in different views.

Filling the spreadsheet by the HIS development team

We had one main contact person who was in charge to make the communication bridge with the development team. We sent him the spreadsheet and the BPMN process models with clear instructions on how to fill the spreadsheet. This procedure was performed with both versions of the spreadsheet. For the second version, we made it clear that the developers could choose any of the two tabs to fill. During the “filling of the spreadsheet” step, we kept direct contact with the development team to solve their doubts. When receiving a newly filled part of the spreadsheet, we immediately reviewed it to check if there was any white cell (cell not filled) and to check if the cells were with a coherent value (e.g. we asked for “prescription of medication” and we received “prescription procedure” in the table name – this seems clearly to be not right). In the case of any problem identified, we contacted them to discuss and update the information.

Extraction of data

Based upon the fields identified in the previous step, SQL queries were written to extract the relevant fields from the HIS. For the extraction, we had to anonymize patient and hospitalization data to guarantee that no-one outside the hospital could identify a patient or link the extracted data with the hospital database. In this stage, to anonymize the data:

1. We encrypted any identification code like patient and health professional codes, chart number, hospitalization number, and prescription and administration ids;
2. Rather than storing the date of birth, we calculated the age of patients according to the admission hospitalization date. Patients older than 90 years old have a higher probability of being identified, thus all of these cases (> 90y) were classified as 90 years old;
3. Patients with a weight greater than 130kg also have a greater probability of identification, thus all of them (> 130kg) were classified as 130kg;
4. All extracted dates were shifted to a given time interval to further remove context which could lead to identification of patient data;
5. For text fields (like clinical notes and discharge summaries) we anonymized names of patients and professionals, specific numbers like chart, hospitalization, telephone, bed numbers.

Results

Regarding the “mapping the process” step (item 2 of the Methods section), several iterations were required to ensure we understood health professionals correctly and vice versa. For us it was a challenge to identify the commonalities and differences in treatment of different patient groups, as defined per severity, age, or list of comorbidities.

The “identification of tables and fields of the database” (item 3) was performed by 10 developers. All of them filled the module oriented tab. Only one developer filled both tabs. At the end we could successfully fill all cells of the spreadsheet. For this step, it was fundamental to have a single contact point to orchestrate the work.

Regarding the “extraction of data” step (item 4), we extracted 4,516 sepsis hospital encounters for a period of two years. We also extracted all (not only sepsis) 61,260 hospital encounters for a period of 2 months, to collect information regarding the workload of professionals (to answer our fourth research question). All the information is present in 57 tables and more than 600 fields.

Discussion

Mapping the sepsis processes was not an easy task, mainly because the communication between two different teams (health and information technology) is very challenging. The use of activity cards (cards filled by the hospital staff containing questions regarding each activity of the process; e.g. step description, notification process, tools used) helped us to understand better the processes. In addition, when validating the processes, the hospital staff could easily understand the BPMN notation (after an explanation of its elements). Thus, the BPMN models were important tools in the communication process between our team and the hospital staff.

Regarding the “identification of tables and fields of the database” step (item 3), all developers filled the module

oriented tab. In our understanding, the process oriented tab was difficult for the development team to work with, since they had to search in all spreadsheet for the fields that they were responsible for. Indeed, for one step a clinical user may have to work in multiple modules, meaning that the attributes would be distributed over the system; and the development team is organized in such a way that sub-teams are responsible for individual modules. We believe that the module oriented tab helped them to easily identify the required fields.

It is important to mention that it was only possible to convert the process oriented view to the module oriented view since we had access to the hospital information system and we had shadowed clinical users during their use of the system. In addition, some of our researchers had previous experience in HIS architecture.

Columns F, G, H and I from the spreadsheet are used to collect the name of the fields regarding the time and user role that executed and registered an action. Professionals that execute actions are not necessarily the professionals who document them. These columns are very important for process mining research and they could lead to interesting insights when it is performed analyses using performance checking or social networks techniques. The executed time and user role might not always be available, and then the best alternative for process mining is to use the information of the registration of data, however taking into account the assumption made in subsequent analyses.

The extraction of non-structured fields is very important for process mining research. These fields may have some information that can be converted to process activities or they can help to find answers for some questions. For example, when a deviation is discovered in a treatment process it is important to understand why certain cases took this unexpected path. Some answers can be found studying clinical notes discovering for example, that patients that followed this deviation were in a critical stage of the disease. Friedrich et al. [11] discussed that it is estimated that 85% of the information of companies is stored in non-structure format and this source of information can be important for creating models. Most of process mining techniques need a very structured event log as input, and the use of natural language processing (NLP) techniques takes an important role for process mining research since it allows to convert unstructured data to concrete events.

Performing this work we could understand that to extract data from a HIS system is not an easy task: a lot of pre-work, knowledge about the treatment process but also about the HIS and its use, and collaboration with multi-disciplinary teams are needed. In general, the complete extraction process took us more time than we expected. Below we present the main reasons:

1. The HIS was not originally designed with the sepsis management pathway, and therefore it was a challenge for us to collect all the information that is relevant for the pathway and to identify where it was located in the system;
2. Limited time availability of the development team to support this initiative;
3. We had to deal with challenges to understanding the data structure of the system;
4. We had to guarantee that all patient and hospitalization information was properly anonymized by our scripts (especially in the non-structured fields);
5. We had limited time to run the scripts to not compromise the use of the system by the hospital.

Our initial attempts of process mining analyses encompassed the evaluation of conformance of a simple sepsis process [12]. It took us great effort to check the quality of the data and to create a simple event log, however this opened up subsequent analyses. Later we started a more complete analysis and we could identify the AS-IS sepsis treatment process in the emergency department, identify deviations, and identify bottlenecks. During this investigation, we had to deal with many challenges (e.g. creating specific events, filtering the right cases, dealing with missing timestamps) and it was very important to have the close participation of health professionals. The results of the research study are very promising. The next step is to validate the results with the hospital staff.

The investigation of healthcare processes is far from a trivial task. Healthcare processes tend to be very flexible and complex. Healthcare professionals play an important role providing health information, analyzing and interpreting results, getting insights and guiding to new analyses. Data scientists know methods and tools to organize and consolidate data in a proper way, and they can propose solutions to optimize resources and improve processes. To reconstruct, analyze and improve healthcare processes it is important to use smart approaches combining informatics/engineering techniques with healthcare knowledge. That means that data scientists and health professionals should work always together in all stages of a research to provide meaningful results. This approach should also be applied in the extraction phase.

We believe that the extraction steps we presented in this work (applying specific adjustments) could be applicable to other healthcare fields of research that uses, for example, data mining, simulation and neural networks techniques.

The extraction phase requires a lot of attention, active and clear communication with external teams, to guarantee that the extracted data will have quality and will allow to perform the research correctly. Getting the wrong data will result in wrong results for the entire research ("garbage in, garbage out" as they say).

Conclusions

With this work, we share our experience in extracting data from a hospital information system to perform process mining research. Any errors made in the extraction phase will have implications on subsequent analyses, thus it is essential to devote a great deal of attention in this phase, even if it is time- and resource-intensive to perform all activities, with the goal of ensuring high quality of the extracted data.

Acknowledgements

The authors thank all professionals from the hospital and the HIS development team for supporting us in the extraction phase.

The authors thank CAPES and Philips Research for funding the development of this work.

References

- [1] W. van der Aalst, *Process Mining: Data Science in Action*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. doi:10.1007/978-3-662-49851-4.
- [2] Process Mining Group - Math&CS department - Eindhoven University of Technology, Process Mining, (2016).

- http://www.processmining.org/ (accessed December 11, 2016).
- [3] A.O. García, D.P. Alfonso, and O.U.L. Armenteros, Analysis of Hospital Processes with Process Mining Techniques, in: *MedInfo*, 2015, pp. 310–314. doi:10.3233/978-1-61499-564-7-310.
- [4] E. Rojas, J. Muñoz-Gama, M. Sepúlveda, and D. Capurro, Process mining in healthcare: A literature review, *J. Biomed. Inform.* **61** (2016), 224–236. doi:10.1016/j.jbi.2016.04.007.
- [5] W. Yang, and Q. Su, Process mining for clinical pathway: Literature review and future directions, in: *2014 11th Int. Conf. Serv. Syst. Manag.*, IEEE, 2014, pp. 1–5. doi:10.1109/ICSSSM.2014.6943412.
- [6] D. Forsberg, B. Rosipko, and J.L. Sunshine, Analyzing PACS Usage Patterns by Means of Process Mining: Steps Toward a More Detailed Workflow Analysis in Radiology, *J. Digit. Imaging.* **29** (2016), 47–58. doi:10.1007/s10278-015-9824-2.
- [7] P. Rattanavayakorn, and W. Premchaiswadi, Analysis of the social network miner (working together) of physicians, in: *2015 13th Int. Conf. ICT Knowl. Eng. (ICT Knowl. Eng. 2015)*, IEEE, 2015, pp. 121–124. doi:10.1109/ICTKE.2015.7368482.
- [8] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, and W. van der Aalst, Process mining techniques: an application to stroke care, *Stud. Health Technol. Inform.* **136** (2008), 573–578. <http://www.ncbi.nlm.nih.gov/pubmed/18487792>.
- [9] Z. Huang, X. Lu, H. Duan, and W. Fan, Summarizing clinical pathways from event logs, *J. Biomed. Inform.* **46** (2013), 111–127.
- [10] R.S. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst, and P.J.M. Bakker, Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital, in: *BIOSTEC*, Springer, 2008, pp. 425–438. doi:10.1007/978-3-540-92219-3_32.
- [11] F. Friedrich, J. Mendling, and F. Puhlmann, Process model generation from natural language text, in: *Int. Conf. Adv. Inf. Syst. Eng.*, 2011, pp. 482–496.
- [12] G.-J. de Vries, R.A.Q. Neira, G. Geleijnse, P. Dixit, and B.F. Mazza, Towards Process Mining of EMR Data - Case Study for Sepsis Management, in: *BIOSTEC*, 2017.

Address for correspondence

Ricardo Alfredo Quintano Neira

e-mail: ricardo.quintano.pucio@philips.com

Development of a Deep Learning Algorithm for Automatic Diagnosis of Diabetic Retinopathy

Manoj Raju, Venkatesh Pagidimarri, Ryan Barreto,
Amrit Kadam, Vamsichandra Kasivajjala, Arun Aswath

Enlightiks Business Solutions Private Limited – a Practo Company, Bangalore, Karnataka, India

Abstract

This paper mainly focuses on the deep learning application in classifying the stage of diabetic retinopathy and detecting the laterality of the eye using fundus images. Diabetic retinopathy is a chronic, progressive, sight-threatening disease of the retinal blood vessels. Ophthalmologists diagnose diabetic retinopathy through early funduscopy screening. Normally, there is a time delay in reporting and intervention, apart from the financial cost and risk of blindness associated with it. Using a convolutional neural network based approach for automatic diagnosis of diabetic retinopathy, we trained the prediction network on the publicly available Kaggle dataset. Approximately 35,000 images were used to train the network, which observed a sensitivity of 80.28% and a specificity of 92.29% on the validation dataset of ~53,000 images. Using 8,810 images, the network was trained for detecting the laterality of the eye and observed an accuracy of 93.28% on the validation set of 8,816 images.

Keywords:

Neural Networks (Computer); Diabetic Retinopathy; [Artificial Intelligence](#).

Introduction

Diabetic Retinopathy (DR) is one of the avoidable causes of blindness. Approximately 33% of people in the USA living with diabetes develop some stage of DR, out of which 10% evolve to a vision-threatening form of the disease [1]. Detecting DR is a time-consuming procedure that requires a trained clinician to evaluate digital funduscopy retinal images. Hence, early and quick diagnosis is a critical aspect in treatment of DR.

Computer Vision (CV) works around building Artificial Intelligence (AI) systems by interpreting information from digital images. Deep learning algorithms have the potential to produce such systems, since they do not revolve around handcrafted features. Convolutional Neural Networks (CNNs), a branch of Deep Learning, is the state of the art algorithm for image classification [7, 9]. From the 1970s and onwards, neural network based approaches were built for pattern recognition. They did not gain prominence due to their ineffectiveness to handle images of a high resolution.

CNN architecture is designed to take advantage of the 2D structure of an input image. It uses two main concepts, namely, local receptive fields and pooling. In CNN, the connections are localized to small regions called local receptive fields, and each input pixel, in turn, is not connected to all the hidden neurons. CNNs also contain pooling layers which help generate translation of invariant features and simplifies the information from the convolutional layer.

Despite having many attractive features, CNNs are still expensive to apply in high-resolution colored images. Current Graphical Processing Units (GPUs), paired with the highly-optimized implementation of 2D convolution are powerful enough to train large datasets of high-resolution images [5].

The hypothesis that CNNs have the potential to improve the efficiency and speed of DR screening, and thereby help prevent visual loss and blindness from this destructive disease, is the basis of this work. The other interesting work is the building of a second convolutional net which detects the laterality of the eye. We were unable to locate any other paper which identifies the laterality of the eye using fundus images along with the detection of the DR stage. Hence, our system is complete in a way that, when multiple images of a patient are uploaded, the system:

- Detects the laterality of the eye (Left vs Right)
- Detects the stage of DR for each eye
- Provides a diagnosis report for each eye

Few research projects have been conducted on the automatic detection of DR, but most of them were carried out with Support Vector Machines (SVMs) or fully connected neural networks. Gardner et al [2] used a neural network to detect diabetic features in fundus images. A fully connected neural network was trained on 147 diabetic and 32 normal fundus eye images. The blood vessels, exudates and hemorrhages were detected with an accuracy of 91.7%, 93.1% and 73.8% respectively.

Usher et al [3] in their work built a system using macula centered color retinal images from 1,273 patients, out of which 500, were used for training the model. Image pre-processing, identification of normal structures and candidate lesions, extraction of candidate lesion features were performed to identify if a patient was normal/abnormal. The trained system was evaluated using the remaining images and observed a sensitivity of 95.1% and a specificity of 46.3%.

Pratt et al [4] in their work managed to use a CNN based approach for automatic detection of DR on the Kaggle dataset (www.kaggle.com) comprising 80,000 fundus eye images. They observed a sensitivity of 30%, specificity of 95% and a 75% accuracy. This is claimed as one of the first papers on classifying DR into five stages.

Google (Gulshan et al [6]), in its vision of deep learning augmented healthcare, has trained a CNN in detecting referable DR using 128,175 training images which were labelled by 54 licensed ophthalmologists and trainees in a massive exercise organized for annotating the dataset. They achieved an average sensitivity of 92.72% and specificity of 95.97%. While the numbers achieved are better here, it requires massive datasets and considerable investments.

This paper provides an overview of the dataset, hardware and software configuration used, architecture of the CNN, the way the model was trained and the observed results with a note on future possibilities.

Methods

DR progresses through four stages, namely: mild, moderate, severe nonproliferative DR (NPDR) and the advanced stages of proliferative DR (PDR). During mild NPDR, small areas of balloon-like swelling in the blood vessels of the retina called microaneurysms occur. As the disease progresses, multiple microaneurysms, hemorrhages, venous beading and cotton wool spots (CWS) occur which make the patients lose their ability to transport blood to the retina, termed as moderate NPDR. The diagnosis is severe NPDR when the abnormalities intensify and are seen on multiple quadrants of the eye and the patient has either of the following: diffuse intra-retinal hemorrhages and microaneurysms in 4 quadrants, venous beading in ≥ 2 quadrants, or intraretinal microvascular abnormalities (IRMA) in ≥ 1 quadrant. As the disease progresses, new blood vessels grow and this stage is termed PDR. The new blood vessels are fragile and are prone to bleeding and cause retinal destruction.

Dataset, Hardware and Software

EyePacs, a telemedicine platform which helps prevent vision impairment from DR (<http://www.eyepacs.com/>), provided the datasets used for training and model validation. EyePacs uses telemedicine to build sustainable DR screening programs in community clinics. It links primary care providers with ophthalmologists, helping early detection of sight-threatening cases of DR. The data was accessible via the Kaggle website. The distribution of each type of image used for training from the dataset is shown in Table 1. A few of the eye images can be seen in Figure 1.

Table 1 - Training Dataset

Type	Number	Percentage
No DR (Normal)	25,810	73.48%
Mild NPDR	2,443	6.96%
Moderate NPDR	5,292	15.07%
Severe NPDR	873	2.48%
Proliferative DR	708	2.01%

Due to the challenges faced during the CNN model training, namely, long training duration and the daunting hyper-parameter tuning, only 35,126 out of 88,702 images are used. The images were of different resolution, cameras, angles, lighting, noise and artifact levels (Figure 2). The biggest challenge the algorithm had to encounter was to generalize all these cases and avoid overfitting.

The images range up to 4500 x 3500 pixels having 3 channels (R, G, B). Processing such huge images on a CNN required a high-end GPU. Two machines were used to train our model.

- NVIDIA GTX980Ti GPU containing 2,816 cores with a video memory of 6GB. The CPU contains an intel core i7 processor with 16GB memory and an Ubuntu OS.
- Amazon EC2 instance containing NVIDIA GPU with 1,536 cores, a video memory of 4GB and an Intel Xeon E5-2670 processor with 15GB memory.

The software used were, Python, NumPy and Theano (www.deeplearning.net/software/theano/), in combination with the cuDNN library. For convenience, we used Lasagne

(www.lasagne.readthedocs.io/), a library built on top of Theano to build and train CNN. Preprocessing of the images was done using the Python Imaging Library (PIL) and the scikit-image was used for the augmentation process.

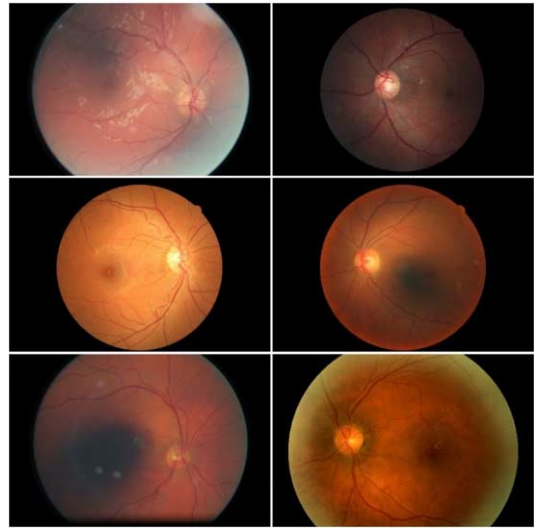


Figure 1 – Sample Fundus Eye Images



Figure 2 – Noisy Fundus Images

Diabetic Retinopathy Prediction

Data Preprocessing

The images were captured using different variants of fundus cameras, each having their unique recording method, due to which the exposure and lighting varied. In addition, images were also too large to be trained with CNN directly. The following pre-processing steps were followed:

- Cropping the image – The images were cropped by distinguishing the foreground from background.
- Resizing the image – The images were resized to 128x128, 256x256, 512x512, 768x768 and 1024x1024 pixels.

- Normalizing the channels – ZMUV (Zero Mean Unit Variance) was performed on each channel.

Data Augmentation

Data augmentation was performed randomly on every image over all epochs. The data augmentation steps performed were:

- Rotation – Images randomly rotated between 0 and 360 degrees
- Translation – Randomly shifted between -40 and 40 pixels
- Zoom – Randomly stretched between (1/1.2, 1.2)
- Flip – Randomly flipping of images
- Color augmentation – Krizhevsky [5] color augmentation
- Image centering

The output size after data augmentation were 112x112, 224x224, 448x448, 672x672, 896x896 pixels for 128x128, 256x256, 512x512, 768x768 and 1024x1024 pixel images respectively.

Network Training

Deep learning is the process of training a neural network to perform a given task [6]. The deep learning algorithm (CNN) computes DR severity by analyzing the pixel intensities over each channel separately. Although the algorithm does not detect the lesions in the eye images explicitly, it learns to recognize them using the filter weights.

The training set comprised of 35,126 images for which the labels were provided by the Kaggle team. The problem was treated as a regression problem (with network output threshold at 0.5, 1.5, 2.5, 3.5) to calculate the accuracy, sensitivity and specificity. The network architecture for DR prediction is shown in Table 2. The network used in this work is inspired by Oxfordnet [7].

Table 2 – Network Architecture for DR Prediction

Layer	No of Filters	Filter Size	Stride	Pad	Output Size
Input					448x448
Convolution	32	4x4	2		224x224
Convolution	32	4x4	1	2	225x225
Maxpool		3x3	2		112x112
Convolution	64	4x4	2		56x56
Convolution	64	4x4	1	2	57x57
Convolution	64	4x4	1		56x56
Maxpool		3x3	2		27x27
Convolution	128	4x4	1	2	28x28
Convolution	128	4x4	1		27x27
Convolution	128	4x4	1	2	28x28
Maxpool		3x3	2		13x13
Convolution	256	4x4	1	2	14x14
Convolution	256	4x4	1		13x13
Convolution	256	4x4	1	2	14x14
Maxpool		3x3	2		6x6
Convolution	512	4x4	1		5x5
RMSpool		3x3	3		2x2
Dropout					
Fully connected	1024				
Feature pool	512				
Dropout					
Fully connected	1024				
Feature pool	512				
Fully connected	1				

- Input - holds the raw pixel values of the image. Here, an image having the width of 448, height of 448, and three color channels R, G, B (depth = 3).
- Convolution – computes a dot product between the filter weights and a small region they are connected to in the input volume.
- Maxpool – performs a down sampling operation along the width and height. A maxpool of 3x3 will obtain the maximum of the local region of size 3x3.
- RMSpool – also performs a down sampling operation. It calculates the root mean square (RMS) value of the local region.
- Dropout – used to prevent overfitting. Nodes were randomly dropped with a probability, $p=0.5$.
- Fully connected – each neuron in this layer will be connected to all the inputs from the previous layer.
- Feature pool – It is an activation function which is the max of the inputs. In our case, we use a pool size of 2.

There are about 12 million learning parameters for the network. The parameters are randomly initialized using orthogonal weight initialization. The mean squared error objective between the known label values and the network output for each image was calculated and in turn used to modify the weighting parameters during a back propagation step to reduce the training loss. This process was repeated for every image in a training set over 250 epochs with data augmentation at each step. The parameters calculated were general enough to work on new fundusoscopic retinal images. The classes in the dataset were highly imbalanced (Table 1) due to which the CNN tends to bias towards the majority class. For training, a resampling of the images was done in a way that all classes were present in equal proportion. The weights of the rare classes were gradually reduced that left us with a final resampling of weights of 1, 2, 2, 2, 2 for normal, mild, moderate, severe NPDR and PDR respectively. The network training started with a batch size of 32, and a learning rate of 0.005. The learning rate was constantly decreased as the number of epochs grew and ended up with a learning rate of 0.00005. A leaky rectified linear unit (ReLU) of 0.01 was used after every convolutional and fully connected layer. An L2 regularization with a factor of 0.0005 was applied on every layer.

For spatial invariance, RMS pooling was used at the last pooling stage of the network. The output of the RMS pool layer was also used as features for the blending network (Table 3) which was trained to improve the prediction results. To increase the quality of the features, feature extraction was repeated up to 50 times with different augmentations per image, and the mean and standard deviation of each feature was used as input to the blending network.

Total Features after RMS pool layer = (Number of filters * Output size after RMS pool) * 2 = (512 * 4) * 2 = 4096

Table 3 – Blending Network

No	Layer	Output size
1	Input	4096
2	Fully connected	32
3	Feature pool	16
4	Fully connected	32
5	Maxout	16
6	Fully connected	1

All features were normalized to have a zero mean and unit variance and were used to train a simple fully connected network. A ReLU of 0.01 was introduced after each fully connected layer and an L2 regularization with a factor of 0.005 was applied to every layer. The batch size used was 128 with a mean squared error objective.

Laterality Detection

Data Preprocessing

Laterality detection prediction used the same preprocessing techniques with uncropped images. The images in the dataset were at times inverted, which needed to be identified to detect the laterality (Figure 3). The inversion of the image is detected by the absence of a notch.

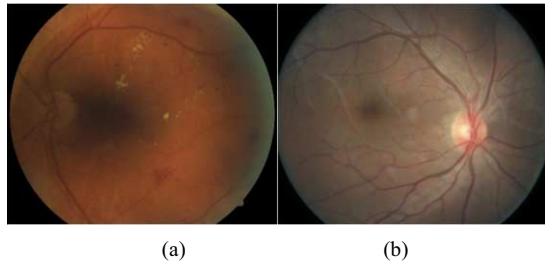


Figure 3 – (a) Non-inverted image (b) Inverted image

Data Augmentation

The same data augmentation process, described for DR prediction was performed, except that the images were not flipped.

Network Training

The training set comprised of 8,810 images for which the labels were provided by the Kaggle team. The laterality detection challenge was treated as a regression problem with a network output threshold at 0.5 to calculate the accuracy.

Table 4 – Laterality Detection Network

Layer	No of Filters	Filter Size	Stride	Pa d	Output Size
Input					448x448
Convolution	32	7x7	1		442x442
Maxpool	32	3x3	2		221x221
Convolution	64	7x7	1		215x215
Convolution	64	7x7	1	2	213x213
Convolution	64	7x7	1		207x207
Maxpool		3x3	2		103x103
Convolution	128	7x7	2		49x49
Convolution	128	7x7	1	2	47x47
Convolution	128	7x7	1		41x41
Maxpool		3x3	2		20x20
Convolution	256	7x7	1	2	18x18
Convolution	256	7x7	1		12x12
Convolution	256	7x7	1	2	10x10
Maxpool		3x3	2		4x4
Convolution	512	7x7	1		4x4
RMSpool		3x3	3		2x2
Dropout					
Fully connected	1024				
Feature pool	512				
Dropout					
Fully connected	1024				
Feature pool	512				
Fully connected	1				

The network architecture for laterality detection is shown in Table 4. The number of learning parameters for this network are about 31 million.

The training process for the network remained similar to the DR prediction network in the weight initialization, loss function definition, regularization technique, data augmentation steps at each epoch and the number of epochs for training. The images used for training contain 50% for the left eye and 50% for the right eye and thus did not require resampling strategies.

A batch size of 32 was used to train the network. The training of the network was started with a learning rate of 0.003 (reduced as the number of epochs grew). Usage of ReLU remained the same as the DR prediction network.

RMS pooling was used for spatial invariance and feature extraction for the blending network (Table 3). The feature extraction was repeated up to 5 times with different augmentations per image and used the mean and standard deviation of each feature as input to our blending network. The total number of features used for this network remained 4096.

Application Overview

An authenticated Java application was built, which accepts patient info (patient name and patient id) and multiple images of the patient and provides a downloadable diagnosis report. The report contains the laterality of the eye, the stage of DR for each image, the follow-up advice and provides an overall diagnosis for each eye.

Results

Image characteristics used for training the network for DR prediction have been summarized in Table 1. Out of the 35,126 images used for training, 9,316 images (26.52%) had DR. The validation set consisted of 53,126 images (Table 5), of which 13,593 images (25.59%) had DR and graded between 0 and 4 (0 –no DR, 1 –mild NPDR, 2 –moderate NPDR, 3 –severe NPDR, 4 –proliferative DR). To train the network, 4x4 filters were used, which observed a specificity of 92.29% and a sensitivity of 80.28%. The confusion matrix of the classification results can be seen in Table 6.

Table 5 – Validation Dataset

Type	Number	Percentage
No DR	39533	73.79%
Mild NPDR	3762	7.02%
Moderate NPDR	7861	14.67%
Severe NPDR	1214	2.27%
Proliferative DR	1206	2.25%

Table 6 – Confusion Matrix (DR Prediction)

		Predicted				
		0	1	2	3	4
Actual	0	36484	2488	472	66	23
	1	1725	1271	753	13	0
	2	1016	1301	3934	1513	97
	3	18	26	232	786	152
	4	10	27	131	402	636

From Table 6, we can see that there are 10 images which have been totally rejected by the system and classified as normal. Also, 23 images which were normal have been classified as proliferative DR. The main cause for these misclassifications are the noise in the images (Fig 4). These images skew the prediction results.

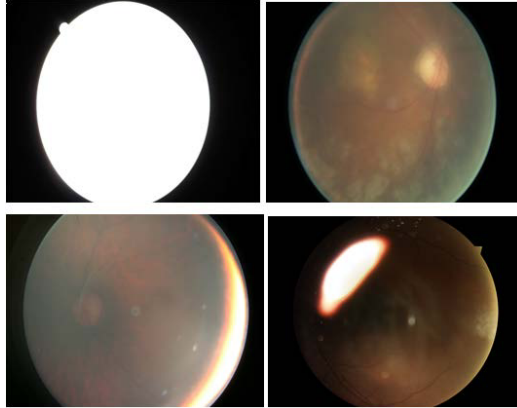


Figure 4 – Noisy Image Labeling

The confusion matrix for laterality detection is shown in Table 7. A total of 8,810 images were used to train the model which contains 4,405 right eye and 4,405 left eye images. The laterality of the eye was detected with an accuracy rate of 93.3% in a validation set containing 4,408 right and 4,408 left images.

Table 7 – Confusion Matrix (Laterality Detection)

		Predicted	
		Left	Right
Actual	Left	4156	252
	Right	340	4068

The exciting aspect of the network is its ability to classify thousands of images every minute which helps in providing a real-time diagnosis for DR.

Discussion

The network was trained using filters of size 3x3, 4x4, 5x5 and 7x7 for DR prediction and laterality detection. The best results were observed with 4x4 filters for DR prediction and 7x7 filters for laterality detection. The smaller filter size was better suited for DR prediction due to the lesions in the images which were of smaller size, such as, microaneurysms. In contrast, laterality detection requires identification of the direction of the blood vessels from the disk for which larger filter sizes suited better. Categorical cross entropy, the objective function used for classification, followed a diverging trend as the epochs progressed. Hence, regression was a better suited option for training the model. The features generated from multiple networks were ensemble but did not observe substantial change in the prediction results. Several image pre-processing strategies, such as edge enhancement, image smoothing, etc. were performed to see if learning could be improved, but neither of them provided better prediction results. Laterality detection was also observed to be inaccurate when the inverted images were mixed with non-inverted

images. A dip in the laterality detection accuracy was also seen for images with no disk and macula.

The classification of Diabetic Macular Edema (DME) is an important addition required in the system. We have collaborated with ophthalmologists in getting DME annotated on the image. Deeper neural networks such as Residual Networks [8] or a variant of GoogLeNet [9] needs to be analyzed on the dataset.

Conclusion

We observed high sensitivity and specificity in the detection of DR staging, as well as laterality of the eye from fundus retinal images, using a CNN based algorithm. This demonstrates the potential of CNNs to automatically classify fundus images based on laterality and severity in real time. Further research is warranted on automatic classification of DME. Other parameters such as patient demographics, past history, family history etc. need to be analyzed to improve the prediction accuracy. Criteria need to be incorporated to filter out images with poor quality. Lesions on fundal images must be identified to help ophthalmologists with their diagnoses. With the incorporation of the improvement areas identified, detection of DR staging and laterality detection of the eye from fundus retinal images would be more robust.

References

- [1] J.W. Yau, S.L. Rogers, R. Kawasaki, E.L. Lamoureux, J.W. Kowalski, T. Bek, S.J. Chen, J.M. Dekker, A. Fletcher, and J. Grauslund, Global prevalence and major risk factors of diabetic retinopathy, *Diabetes Care* **35** (2012), 556-564.
- [2] G.G. Gardner, D. Keating, T.H. Williamson, and A.T. Elliott, Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, *British Journal of Ophthalmology* **80** (1996), 940-944.
- [3] D. Usher, M. Dumskyj, M. Himaga, T.H. Williamson, S. Nussey, and J. Boyce, Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening, *Diabetic Medicine* **21** (2004), 84-90.
- [4] H. Pratt, F. Coenen, D.M. Broadbent, S.P. Harding, and Y. Zheng, Convolutional Neural Networks for Diabetic Retinopathy, *Procedia Computer Science* **90** (2016), 200-205.
- [5] A. Krizhevsky, I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [6] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams and J. Cuadros, Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, *JAMA* **316** (2016), 2402.
- [7] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
- [8] K. He, X. Zhang, S. Ren and Jian Sun. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015).
- [9] C. Szegedy, L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, in: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2015.

Addresses for Correspondence

Manoj Raju manoj.raju@enlightiks.com

Venkatesh Pagidimarri venkatesh.pagidimarri@enlightiks.com

Envisioning Precision Healthcare Informatics: A Unified Framework

Arkalgud Ramaprasad^a, Nanda Kumar Bidare Sastry^b, Thant Syn^c

^a Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL, USA

^b M S Ramaiah College and Hospitals, Bengaluru, India

^c Division of International Business and Technology Studies, Texas A&M International University, Laredo, TX, USA

Abstract

We present a unified framework for envisioning precision healthcare informatics (PHI). The framework, presented as a high-level ontology, deconstructs PHI into six dimensions derived from the informatics and healthcare disciplines—the structure, function, and semiotics of informatics, and the stakeholders, care, and value of precision healthcare. Each dimension is articulated as a taxonomy of its constituent elements. Thus, the framework encapsulates the combinatorial complexity of PHI and can be used to describe its 37,800 potential components in natural English. The framework will help envision PHI systematically, systemically, and symmetrically in contrast to the often simplified, selective, and skewed approaches. The framework can be used to map the state-of-the-research, -practice, and -policy on PHI, discover the gaps within and between the states, and develop a roadmap for the future. It can be scaled and extended as PHI evolves by adding dimensions, and extending/refining the taxonomies.

Keywords:

Precision Medicine; Medical Informatics

Introduction

Informatics will be key to realizing the full potential of precision medicine and solving the problems likely to emerge from it [1]. Many authors have highlighted the characteristics of informatics required for the purpose [2-7]. However, there is no unified framework to envision precision healthcare informatics (PHI). We need to make the PHI ‘elephant’ visible like the allegorical wise sage does in the story of the six blind men and the elephant.

We present a unified framework for envisioning PHI. We emphasize the indefinite article ‘a’ instead of the definite article ‘the’ in characterizing the framework. In doing so, we recognize the possibility of other equally plausible frameworks developed from other perspectives, although there is none in the literature now. It draws upon many simple and selective frameworks in the literature, some explicitly articulated and others implicitly incorporated. The framework is both systemic in its coverage and systematic in its development. Its outlook is symmetrically neutral with respect to the value of precision healthcare—recognizing that the value could be both positive and negative. It is presented using structured natural English and can be easily understood, interpreted, and applied by the researchers, policy makers and practitioners. In the following, we describe the framework,

discuss its application, and conclude with a roadmap for future research on the topic.

Methods

PHI is a complex domain. The challenge of constructing a unified framework for designing, developing, and applying informatics to precision healthcare is to construct one which is logical, parsimonious, and complete. It must be logical in the deconstruction of the domain, and parsimonious yet complete in the representation of the domain. It must be a closed description of PHI in its entirety yet adaptable to the evolution of the domain. We represent the combinatorial complexity of PHI using an ontology.

The ontology represents our conceptualization of the PHI domain [8]. It is an “explicit specification of [our] conceptualization,” [9] and can be used to systematize the description of the complexity of PHI domain knowledge [10]. The ontology organizes the terminologies and taxonomies of the PHI domain. “Our acceptance of [the] ontology is... similar in principle to our acceptance of a scientific theory, say a system of physics; we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged.” [11] It is a domain ontology that “helps identify the semantic categories that are involved in understanding discourse in that domain.” [12] Ontologies are used in computer science, medicine, and philosophy. Our ontology of PHI—unified framework—is less formal than computer scientists’, more parsimonious than medical terminologists’, and more pragmatic than philosophers’. It is designed to be actionable and practical, and not abstract and meta-physical. Its granularity matches that of the discourse in research, policy, and practice, and facilitates the mapping and translation of the domain-text to the framework and the framework to the domain-text. It is also adaptable. In the concluding section of the paper we will discuss how the framework can be scaled and zoomed (in and out) to adapt to changing requirements.

The framework is shown in Figure 1. Three illustrative components derived from the framework are shown below it with an example of each. The elements of the framework are defined in the glossary in Appendix A. In the following subsections, we describe the construction of the unified framework and argue about its validity. In the next section, we will articulate the application of the framework.

Precision Healthcare Informatics					
Informatics			Precision Healthcare		
Structure	Function	Semiotics	Stakeholders	Care	Value
Architecture	Acquire	Data	Recipients	Avoidance	Life
Infrastructure	Store	Health Records	Individuals	Prevention	Psychological
Systems	Retrieve	Knowledge	Families	Treatment	Economic
Services	Process		Communities	Cure	Social
Policies	Distribute		Societies	Elimination	Cultural
Processes	Delete		Providers		Ethical
People			Payers		
			Researchers		
			Pharmaceuticals		
			Governments		
			Regulators		

Figure 1 – Unified Framework of Precision Healthcare Informatics

Illustrative Components:

Architecture to acquire data of recipients (individuals) for personalized avoidance of illness for life value of healthcare.

Example: Informatics architecture to systematically obtain data about individuals with a genetic propensity for diabetes to avoid its occurrence and prolong their life.

Processes to distribute knowledge to providers for personalized cure of illness for economic value of healthcare.

Example: Processes to distribute knowledge about emergent, less expensive treatment of an illness.

Policies to delete health records by payers for personalized prevention of illness for ethical value.

Example: Ethical prohibition on insurers from collating healthcare data on families.

Construction of the Framework

The domain of interest is PHI. It is a composite of two subdomains—informatics and precision healthcare. Each has a considerable corpus of research, policies, and practice anchored in very different academic and practice disciplines. Informatics is based primarily on computer and social sciences; precision healthcare draws from a range of medical, biological, physical, social, and other sciences. The latter also draws from the humanities and ethics. The unified framework provides an interdisciplinary perspective to address the problem of PHI. Thus:

PHI = Informatics + Precision Healthcare

Informatics

Informatics can be defined by its Structure, Function, and Semiotics—a commonly used deconstruction. Thus:

Informatics = Structure + Function + Semiotics

There are many ways of defining the Structure of an information system. The traditional taxonomy is: hardware, software, networks, people, policies, and procedures. We have modified the taxonomy for two reasons. First, the emergence of technologies like cloud computing and IoT (internet of things) [13] have blurred the lines between hardware, software, and networks. They constitute the Infrastructure of PHI. Second, there is likely to be multiple information systems competing and collaborating with each other in precision healthcare. They need an Architecture to work effectively. [Note: Elements of the framework in the text are capitalized.] The Systems includes the technical and social systems constituting PHI. The Services are those provided by PHI; the Policies and Processes govern the operation of PHI; and the People are the users and operators of PHI. Thus:

Structure \subset (Architecture, Infrastructure, Systems, Services, Policies, Processes, People)

The traditional Functions of an information system are to Acquire, Store, Retrieve, Process, and Distribute. To this taxonomy, we have added Delete—a hotly debated issue often discussed in relation to the “right to be forgotten” [14]. It is an issue which has become critical in the context of healthcare information privacy and security [15]. Usually these functions are sequential and iterative. Thus:

Function \subset (Acquire, Store, Retrieve, Process, Distribute, Delete)

Semiotics is the repetitive cycle of generation and application of knowledge. The taxonomy of the Semiotics dimension is commonly used [16]. They correspond roughly to morphologies (data), syntactics (information), semantics and pragmatics (knowledge) [16]. Sometimes Wisdom is included as a fourth element of the taxonomy. It is still an ephemeral construct and hence not included in the present taxonomy. They are generally sequential and iterative. Data is translated into Information and Knowledge in the generation phase, and from Knowledge to Data in the application phase.

Semiotics \subset (Data, Information, Knowledge)

Precision Healthcare

There is no universally accepted definition of precision healthcare. It is sometimes used synonymously with personalized healthcare. At the core of these two constructs is the recognition of individual differences in the manifestation of illness, wellness, care of illness, and continuation of wellness. Thus, precision healthcare is the personalization of care to an individual based on his or her characteristics, instead of on the basis of the person’s inherited population characteristics [17]. The reference population may be based on one or many enduring traits of an individual such as genes, age, gender, race, ethnicity, geographical region, and culture. They may also be based on transient states of an individual such as psychological, economic, physical, and social state. Thus, the characteristics considered for precisely targeting healthcare may be a person’s genetic, physical, psychological, cultural, economic, social, and other characteristics. Although the term is frequently used to describe the use of genetic characteristics for personalizing healthcare, its scope is far broader. At its core, it recognizes the importance of simultaneously considering a person’s similarity with a population and his/her differences with the population [17; 18].

Precision medicine is far more semiotics intensive than generalized medicine. Data, information, and knowledge have to be generated and applied at a far finer level of granularity. The structure and function of informatics have to be designed to support the scaling of requirements, qualitatively and quantitatively, by orders of magnitude. Even if the dominant focus is on genetic precision, the care’s life, psychological, economic, social, cultural, and ethical values will determine the efficacy of care. While approach to health and disease is based on population evidence, treatment is often at the individual level requiring a multitude of factors to consider in ensuring a favorable outcome [19].

Further, the precision care could be delivered at different stages of illness (or its absence). It could be to avoid, prevent, treat, cure, or eliminate the illness [20]. The value profile and the informatics requirement for each stage would be different.

There are many stakeholders in personalized healthcare—as in any healthcare. We can broadly classify them as recipients,

providers, payers, researchers, pharmaceuticals, governments, and regulators. The recipients could be further classified as individuals, families, communities, and societies. The stakeholders' perspective on different type of care and their value can vary significantly. Consequently, their requirement of informatics too will vary significantly.

Thus, we can deconstruct precision healthcare as follows:

Precision Healthcare = Stakeholder + Care + Value

Stakeholder \subset (Recipients, Providers, Payers, Researchers, Pharmaceuticals, Governments, Regulators)

Recipients \subset (Individuals, Families, Communities, Societies)

Care \subset (Avoidance, Treatment, Cure, Elimination)

Value \subset (Life, Psychological, Economic, Social, Cultural, Ethical)

The dimensions of the framework are arranged left to right with adjacent words/connectors such that the concatenation of an element from each dimension with adjacent words/connectors creates a natural English sentence illustrating a potential component of PHI. The components and fragments (incomplete components) define the domain.

At the most detailed level, the framework encapsulates 37,800 ($7*6*3*10*5*6$) potential components of PHI. It encapsulates the 'big picture' of the domain and helps visualize the combinatorial complexity of PHI. Three illustrative components and examples are listed in Figure 1. A component may be instantiated in many ways in research, policy, and practice. It may also be instantiated fully or partially (as a fragment). Conversely, a research, policy, or practice may be mapped to a component, a fragment, multiple components, and/or multiple fragments.

The validity of the framework in Figure 1 will determine how well it captures and represents PHI. We draw upon the traditional constructs of validity commonly used in social sciences [21] to justify the face, content, semantic, and systemic validity of the framework of PHI similar to justifying the validity of mHealth ontological framework [22].

Discussion

The proposed framework is a lens to study the anatomy of PHI. For a complex domain like this, there may be other lenses to study the same and each can be encapsulated by a different framework. They will provide different perspectives. We will discuss the present one in detail.

We have discussed the individual dimensions (columns) and elements of the framework while describing the construction of the ontology. Multiple elements of a dimension may coexist independently but may also interact with each other. Thus, many values, types of care, and stakeholders may coexist and interact with each other. Knowing the independent and interacting elements is critical to understanding PHI. The framework can help systematically study and manage the elements' interdependence and interactions. In the following, we discuss how the ontology can be used to systematically study the interaction of: (a) elements within a dimension, (b) elements across two dimensions, and (c) elements across multiple dimensions, to understand the anatomy of PHI at different levels of granularity and complexity.

Combinations Within a Dimension

All possible first-order interactions among the elements of a dimension can be mapped into a table of the dimension with itself. Such a mapping can reveal strong interactions (both constructive and obstructive), weak ones, absent ones, and unexpected ones among the elements. It can also highlight the direction of the interaction—one-way (a to b OR b to a), and two-way (a to b AND b to a). In the following, we will discuss some possible insights from such a mapping of each dimension.

Value

The value of life is central to healthcare and the physicians are sworn to preserve life by their Hippocratic oath. Yet, with medical advances that have led to precision healthcare, one is compelled to consider the other values. How does it affect the recipients' psychological quality of life? Their economic wellbeing? How is it viewed in their social and cultural milieu especially, for example, in the context of terminal illnesses and views about death and dying. Last, the ethics of all of the above. Thus, although some of the values can be identified separately they interact with each other strongly. The informatics for precision healthcare should provide the data, health records, and knowledge to understand and manage the multiple values.

Care

The Care elements are quasi-sequential. One seeks to prevent a condition that cannot be avoided; to treat the one that cannot be prevented; to cure what is treated; and to eliminate what has been cured. Looked at differently, the precision healthcare strategy for a stage of care will depend on the strategies for the prior stages, and will affect those for the subsequent stages. However, sometimes a stage or two may be skipped. Thus, the management of semiotics (data, healthcare records, and knowledge) for care has to be cumulative and supported by the appropriate informatics structure and function.

Stakeholders

The stakeholders act independently and interactively. Their interaction may be cooperative and conflicting. An interaction matrix of the stakeholders will help map the dynamics which can affect the value of PHI. A frequent source of conflict among stakeholders may be the absence of a common base of data, information, and knowledge about precision healthcare. PHI can partially help reconcile the differences by developing a common repository on a universally accessible platform. Such a platform can make the semiotics transparent, even if it does on reconcile the differences.

Semiotics

Data, Information, and Knowledge are bi-directionally sequential. From Data to Knowledge during generation, and in the opposite direction during application. Precision healthcare will require repeated cycles of generation and application of knowledge. Errors in data can lead to errors in knowledge and subsequent errors in actions resulting from the application of that knowledge. The errors may be perpetuated and amplified in subsequent cycles unless they are identified and attenuated. Similarly, valid data can lead to valid knowledge and appropriate action. These too may be perpetuated and amplified in subsequent cycles unless they are overlooked and attenuated. A challenge in the design of PHI is to attenuate the dysfunctional cycles and amplify the functional ones.

Function

The Function elements are strictly sequential—Storage follows Acquisition, Retrieval follows Storage, and so on. They are also repetitive. Consequently, the impact of PHI choices will be propagated from Acquisition to Deletion. For example, inclusion of data on Communities as Recipients in the specification of the system can help acquire data about it and process the same. On the other hand, ignoring community data will result in the inability to personalize healthcare to the community based on information about and knowledge of variations in health in the community. The effects of inclusion/exclusion will be propagated to the subsequent cycles, unless corrected. The inclusion/exclusion of the right data is as likely to be propagated as the inclusion/exclusion of the wrong data.

Structure

The Structure elements are generally hierarchical. Architecture defines the Infrastructure, Infrastructure defines the Systems, and so on. As such the lower-level elements are likely to inherit their properties from the higher ones. Although possible, it is less likely that the properties of the lower elements will be propagated to the higher ones. Thus, elements of PHI which are part of the Architecture will likely permeate the Policies, Processes, and People. However, elements of PHI inserted at the lower levels in Policies, Processes, and People are less likely to be propagated into the Architecture and Infrastructure; in fact, such propagation is likely to be resisted.

Combinations between Dimensions

In addition to interactions among the elements of a dimension, all possible first-order interactions among the elements of a pair of dimensions can be mapped into a table. Such a mapping can reveal strong interactions (both constructive and obstructive), weak ones, absent ones, and unexpected ones between the elements of the two dimensions. It can also reveal the direction of the interaction—one-way (a to b OR b to a) and two-way (a to b AND b to a). With the six dimensions of the framework there are fifteen possible pairs. In the examples in the previous section we have some possible interactions between many of the dimensions.

Conclusion

The proposed framework fills an important gap in the literature on PHI. It defines the domain of PHI systematically, systemically, and symmetrically (by being effect-neutral and hence including both positive and negative effects of PHI). All researchers, policy makers, and practitioners may not universally agree with the framework even though a systematic attempt has been made to include all the key elements from the extant literature. The disagreements can be accommodated by extending the taxonomies by including overlooked elements, reducing them by eliminating redundant elements, coarsening them by combining elements, and refining them by dividing the elements.

The scope of the framework can be broadened by adding absent dimensions and narrowed by deleting present dimensions. For example, a Temporal dimension may be added with three elements—Short, Medium, and Long term as a dimension of Precision Healthcare. Inclusion of this dimension will compel consideration of the Care x Value combinations in each of the time horizons by the stakeholders, rather than in the aggregate. It would also require information

system support with finer granularity, lead times, and response time. On the other hand, the Semiotics dimension may be eliminated if Data, Information, and Knowledge for PHI are aggregated and labeled Information. These properties of scalability (extensibility, reducibility), and zoomability (up and down) make the unified framework versatile to study the domain.

The framework of PHI presented in this paper makes visible the combinatorial complexity of an important and timely topic in healthcare informatics. The ontology is logically constructed but grounded in the theories prevalent in the domain and relevant disciplines. The dimensions are logically specified and not empirically generated. They are deduced from the definition of the domain. Thus, it helps us address the problem of PHI in its entirety rather than fragmentarily. A common tendency in research into a complex domain is to highlight the complexity of the whole domain and then to address selected parts of the domain, hoping that somehow the parts will fit together and enlighten the whole problem. Unfortunately, very often, even after an extensive body of research is accumulated the problem is not illuminated in its entirety. The proposed framework and the mapping of the literature will help obviate the problem.

The framework is a lens to study PHI. There may be other lenses to study the same. Each can be described by an ontology. Each ontology is constructed from the theory and practice of the domain. The logical construction of our ontology minimizes the errors of omission and commission. For example, inclusion of the Stakeholder dimension compels the researchers to explicitly consider different stakeholders. Without consideration for the Stakeholder dimension (error of omission), researchers will be unlikely to advance the field of PHI. By the same token, it may specify PHI (error of commission) stakeholders (for example, Societies) not relevant in a particular context and which should be removed.

Last, the framework is a multi-disciplinary lens. The Structure, Function, and Semiotics dimensions are drawn from the information systems and knowledge management literature; the Stakeholders, Care, and Value are from the healthcare management and related literature. The framework compels the user to analyze the issues surrounding PHI and synthesize solutions by drawing upon these disciplines. One may construct other frameworks/lenses and study the subject from a different perspective.

The framework of PHI can advance the state-of-the-research in the domain. It can be used to systematically identify the 'bright', 'light', and 'blind/blank' spots in the research on the topic. The state-of-the-research can be analyzed by mapping the corpus of research onto the framework. The mapping can be done by identifying categories within each dimension articles in the corpus fit into. An article may be mapped to one or more categories in one or more dimensions, or none. Such mapping will highlight the bright, light, and blind/blank spots in PHI research. The map will help researchers visualize the landscape of the PHI enabling them to set appropriate research direction. Similarly, the state-of-the-practice can be analyzed by mapping current practices onto the framework. The resultant mapping will identify the focus of and gaps in the inclusive practices. The comparison of the ontological maps of research and practice will bring to fore the alignment between the two states, or lack thereof. The gaps between the two states should inform researcher, practitioners, and policy makers alike the need for action to advance research and policy. In other words, the maps can be used as a roadmap for PHI research and practice. It can illuminate the 'big picture' of the domain.

The landscape of PHI will change over time with emerging technologies. The ontology-based road map can be amended to reflect the changing landscape. New categories and dimensions can be added, obsolete ones discarded, and existing ones modified. Changes can also be introduced by the shifting focus in the domain. The finer levels of dimensions and elements can be added to the framework to reflect the greater focus on certain dimensions or categories. On the other hand, sub-categories and -dimensions can be collapsed to echo their diminishing importance in the domain. The shifting focus and direction of research and development can be monitored by analyzing the snapshots of the ontological maps over time. The unified framework can help visualize the past and present of the domain, and envisage its future.

“Precision medicine has the potential to profoundly improve the practice of medicine. However, ... fundamental changes are needed in the infrastructure and mechanisms for data collection, storage and sharing. This will create a continuously learning health-care system with seamless cycling between clinical care and research. ... The building blocks for such a system are already forming and they will accelerate the adoption of precision medicine.” [23] The proposed unified framework for PHI will help envision the roadmap for achieving precision healthcare’s full potential.

References

- [1] J.L. Jameson and D.L. Longo, Precision Medicine—Personalized, Problematic, and Promising, *Obstetrical & Gynecological Survey* **70** (2015), 612-614.
- [2] P.H. Carney, Information Technology and Precision Medicine, *Seminars in Oncology Nursing* **30** (2014), 124-129.
- [3] L.J. Frey, E.V. Bernstam, and J.C. Denny, Precision medicine informatics, *Journal of the American Medical Informatics Association* (2016).
- [4] T.A. Peterson, E. Doughty, and M.G. Kann, Towards Precision Medicine: Advances in Computational Approaches for the Analysis of Human Variants, *Journal of Molecular Biology* **425** (2013), 4047-4063.
- [5] A. Stoner and O. Elemento, A primer on precision medicine informatics, *Briefings in Bioinformatics* (2015).
- [6] J.D. Tenenbaum, P. Avillach, M. Benham-Hutchins, M.K. Breitenstein, E.L. Crowgey, M.A. Hoffman, X. Jiang, S. Madhavan, J.E. Mattison, R. Nagarajan, B. Ray, D. Shin, S. Visweswaran, Z. Zhao, and R.R. Freimuth, An informatics research agenda to support precision medicine: seven key areas, *Journal of the American Medical Informatics Association* **23** (2016), 791-795.
- [7] Y. Zhang, Q. Zhu, and H. Liu, Next generation informatics for big data in precision medicine era, *BioData Mining* **8** (2015), 34.
- [8] T.R. Gruber, Ontology, in: *Encyclopedia of Database Systems*, L. Liu and M.T. Özsu, eds., Springer-Verlag, New York, 2008.
- [9] T.R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *International Journal Human-Computer Studies* **43** (1995), 907-928.
- [10] J.J. Cimino, In defense of the Desiderata, *Journal of Biomedical Informatics* **39** (2006), 299-306.
- [11] W.V.O. Quine, *From a Logical Point of View*, Harvard University Press, Boston, MA, USA, 1961.
- [12] B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins, What Are Ontologies, and Why Do We Need Them?, *IEEE Intelligent Systems* **14** (1999), 20-26.
- [13] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, Internet of Things (IoT): A vision, architectural elements, and future directions, *Future Generation Computer Systems* **29** (2013), 1645-1660.
- [14] J. Rosen, The Right to Be Forgotten, *Stanford Law Review Online* **64** (2012), 88-92.
- [15] E.A. Shoor, Narrowing the Right to Be Forgotten: Why the European Union Needs to Amend the Proposed Data Protection Regulation, *Brooklyn Journal of International Law* **39** (2014), 488-519.
- [16] A. Ramaprasad and A. Rai, Envisioning Management of Information, *Omega-International Journal of Management Science* **24** (1996), 179-193.
- [17] J. Schaffhausen, What Precisely Is Precision Medicine?, *Trends in Pharmacological Sciences* (Forthcoming).
- [18] B. Prasad, M. Vrana, A. Mehrotra, K. Johnson, and D.K. Bhatt, The Promises of Quantitative Proteomics in Precision Medicine, *Journal of Pharmaceutical Sciences* (Forthcoming).
- [19] J.D. Hinman, N.S. Rost, T.W. Leung, J. Montaner, K.W. Muir, S. Brown, J.F. Arenillas, E. Feldmann, and D.S. Liebeskind, Principles of precision medicine in stroke, *Journal of Neurology, Neurosurgery & Psychiatry* (2016).
- [20] L. Zhuang and M.T. Johnson, How Precisely Can Prostate Cancer Be Managed?, *International Neurology Journal* **20** (2016), S120-130.
- [21] L. Brennan, J. Voros, and E. Brady, Paradigms at play and implications for validity in social marketing research, *Journal of Social Marketing* **1** (2011), 100-119.
- [22] J.D. Cameron, A. Ramaprasad, and T. Syn, An Ontology of and Roadmap for mHealth Research, *International Journal of Medical Informatics* **100** (2017), 16-25.
- [23] S.J. Aronson and H.L. Rehm, Building the foundation for genomics in precision medicine, *Nature* **526** (2015), 336-342.

Address for correspondence

Thant Syn, Division of International Business and Technology Studies, A. R. Sanchez, Jr. School of Business, Texas A&M International University, 5201 University Boulevard, Laredo, Texas 78045, USA

Email: thant.syn@tamiu.edu

Tel: +1 (956) 326-2527, Fax: +1 (956) 326-2494.

A Virtual Earth Model of the Dementias in China

Hamish Robertson^a, Nick Nicholas^b, Jo Travaglia^c, Andrew Hayen^c, Andrew Georgiou^d

^a Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia

^b The Demographers Workshop, Sydney, NSW, Australia

^c Faculty of Health Sciences, University of Technology, Sydney, NSW, Australia

^d Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW, Australia

Abstract

This developmental project was undertaken to explore how applying spatial science analysis and visualisation methods might inform societies undergoing significant structural and demographic change. China is rapidly transitioning to an aged society. It already exceeds all other countries in its population aged 65 years and over. Dementia is closely correlated with ageing and intersects with a variety of physical and cognitive disabilities. Information dashboards are a growing part of health and social policy data environments. These visual data applications increasingly include mapping capabilities. In this paper, we explore the utility of a geographic modelling approach to exploring the complex nature of population ageing and the dementias in China.

Keywords:

Geographic Mapping; Ageing; Dementia

Introduction

China's population is ageing rapidly. Current estimates indicate an increase in total population from 1.304 billion people in 2010 to 1.402 billion in 2020 (<https://populationpyramid.net/china/>). In 2015 China's 65+ population overtook the entire population of Japan, which is still the most aged society on earth. And further to this, China's ageing is expected to accelerate from around 2025 as both the overall 65+ group increase and the more dependent oldest old cohorts, aged 80 and over, grow faster still [1]. Average life expectancies increased from just under 72 years in 2000 to almost 76 in 2014 in what is a continuing upward trend (<http://data.worldbank.org/>). China may even exceed Japan as the most aged society by 2030. Figure 1 depicts the rapid expansion and changing composition of China's ageing population. In particular, by 2050 almost one third of the 65+ population will be 80 years and over. This means that the 'old-old' will be a major feature of China's population ageing, as they are in Japan and many European countries. This is a positive outcome for the population given that it reflects rising life expectancies for a huge group of people (120 million by 2050). However, this group is also much more likely to experience a dementia in association with a variety of other age-related health problems, meaning that the prevalence and incidence of sub-acute and acute will conditions rise proportionally.

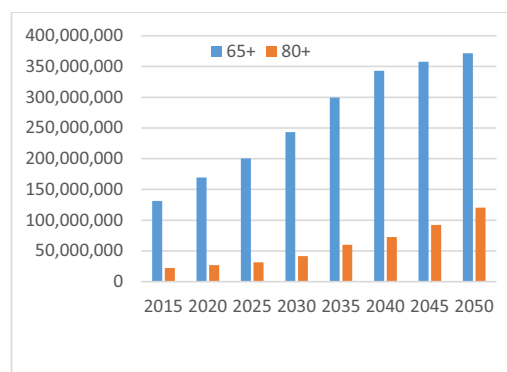


Figure 1- Projected Population Ageing in China, 65+ and 80+ Cohorts. Source: UN Population Projections, 2015 Revision – Medium Variant

Population ageing is generally associated with rising rates of chronic disease and disabling conditions – including both physical and cognitive disabilities. As a result, China can expect rapid expansion in the need for age-related care of all kinds. The dementias, in particular, are likely to prove very challenging because they impair cognitive capacities and intersect with other acute and chronic health problems in older people. In addition, China is a very large country with major structural and resource differences between urban and rural areas. In other words, the geography of ageing in China will be an important consideration in responding to these pressing demographic trends.

This project takes these issues as a starting point and examines them from a spatial health informatics perspective. Health informatics is a growing sector for research and investment globally in a process driven by demographic changes, urbanisation and systemic developments. One of the important roles of health informatics is in informing governments, funders and health providers about how such systems are performing as well as where, how and why variations emerge. In countries with very large populations, such as China, effective health informatics requires a spatial perspective in order to respond adequately and appropriately across distances and at different scales (e.g. provincial, municipal, county, village).

This paper explores one of the options for visualising the complexities of demographic change in equally complex geographic and systemic environments. The value in this approach lies in its capacity to inform a broad audience about growing complexities in an explicitly visual environment. The shift to data visualisation is already an important process in

other industries and we can see how health, aged care and disability – all complex social informatics domains – will benefit from the application of visualisation methods. This is one example of such an information strategy.

Methods

We used age and sex data for all Chinese counties (2872) from the 2010 Chinese census for our demographic modelling. The data were extracted in csv. file formats and manipulated in Microsoft Excel. The population data were sourced from the China Data Center online project at the University of Michigan (<http://chinadatacenter.org/>) through our institutional library subscription. We applied dementia rates to selected cohorts (aged 55+), by age and sex, of these data from the results of the study by Chan et al (2013) to this county population modelling. These rates only included the prevalence rates for the 85-89 age group as the top estimate. This was because the Chinese Census data was only available at the aggregate level for 85 years and above (see Limitations below). The results of this analysis were then mapped onto China's county-level geography using standard geocoding methods to link the data to a latitude and longitude.

We initially mapped these results using a geographic information systems (GIS) package (Maptitude™). The results were then visualized in a secondary analysis stage using Google Earth™ to create a Keyhole Markup Language (KML) format file of the data. This is a method that many GIS packages support internally (i.e. export to KML or Google Earth format) or users can use a free software application like GEGraph (<http://www.sgrillo.net/googleearth/gegraph.htm>). The growing development of three-dimensional mapping is seeing a much closer integration between traditional GIS, and mapping functionality, and these emerging data visualisation methods.

Virtual earth technologies are a growing area of application in China in addition to many other countries, and thus congruent with current research and planning applications [2]. This later stage produced a more dynamic visual information in a three-dimensional (3D) format with which users can engage. The focus here was on applying contemporary spatial methods to health-related information and its visualization. Another consideration was the use of non-proprietary software options to enhance potential accessibility of the results. We note that this is an area undergoing rapid developments and other options are emerging at regular intervals. However, Google Earth™ has the added utility of high recognition outside of the spatial sciences, making it a useful introduction to potential applications in the health and aged care sectors.

Results

The first stage in the analysis developed a detailed estimation of dementia prevalence by age and sex cohorts in the 2010 population at the county level for the whole of China. This method produced a dementia-affected population of just over 10.2 million people, higher than the Chan et al [3] estimate (9.2 million people) but probably lower than actual numbers due to the cut-off estimation at 55 years and over, our inability to access detailed cohorts for the over 85 groups, and the use of a combined rate for males and females.

These results are therefore indicative only and designed to illustrate the relationship between population ageing, geography and a progressive and innovative health information strategy for China – an area which requires further development. It is possible, for example, that

prevalence and incidence rates may differ between urban and rural areas. Chan et al have suggested this is not a major differentiator in their meta-analysis but they also note the limitations of the studies they utilised and suggest more detailed research is warranted.

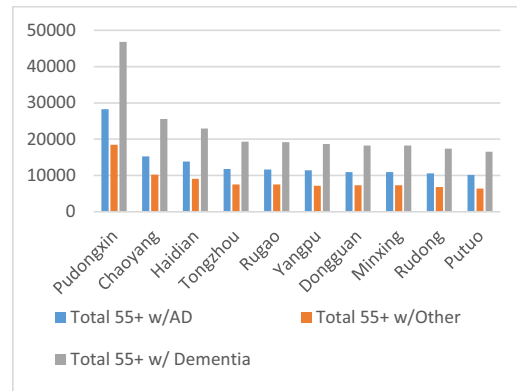


Figure 2: Top 10 Counties by Major Dementia Categories

The chart above shows that the top ten counties in China with the largest dementia population estimates account for less than 250,000 patients (222,869). Of these, 134,917 are likely to have Alzheimer's disease while the remainder have vascular or other dementias. This might seem an unusually low impact weighting for a top ten listing but there are another 2,862 counties, many with similar population sizes and demographic compositions. This also indicates that the geographic impact of the dementias in China is not restricted to specific regions but is a nation-wide phenomenon, even taking into account the massive rural to urban migrations of the past several decades [4].

China's population is ageing and there is a distinct geographical pattern to this process. The illustrations below were generated using the data we modelled in conjunction with the GEGraph software package and Google Earth™. The result is a Keyhole Markup Language (KML) file including dementia estimate data for all counties in China. The data range has been scaled down to reduce the differences between the highest and lowest county data. The image below is a JPEG image but the output is an active file that any user can engage with if they have Google Earth™ installed on a device – desktop, portable device or tablet.

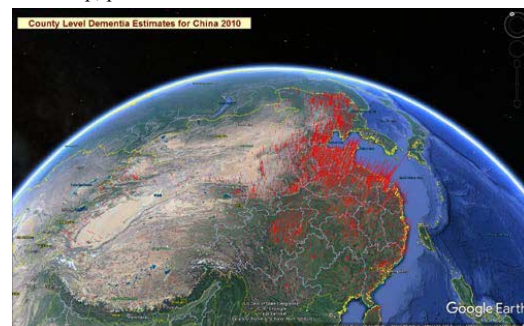


Figure 3: County Level Estimates of Dementia Prevalence in China 2010

The illustration above, Figure 3, shows the results for county-level dementia prevalence in 2010 for the whole of China. This illustrates not only the population distribution by county but also emphasises the eastern coastal and north-eastern

regional density patterning of China's ageing population. Each county's data is illustrated in red for easier visibility and the actual KML file is fully accessible in Google Earth™ which greatly enhances its visual accessibility compared to this reduced detail JPEG image.



Figure 4: County Level Estimates of Alzheimer's Disease in China 2010

In the second Google Earth™ image above, we have segmented out the data estimates for Alzheimer's Disease (AD) specifically, and visualised these in green to avoid any confusion between the different images. As with a geographic information system (GIS) these KML files can be supplemented with files of other data, such as hospital locations or pharmacies, to illustrate and explore the relationships between a specific health condition and health services infrastructure. As health systems and health informatic progress globally, we suggest that these types of additions to the health informatician's toolkit will greatly enhance the utility of health information datasets and health informatics modelling practices. All health and disease problems exhibit a spatial dimension and this type of visualisation can support understanding of these problems and our efforts to manage and resolve them.

In a third Google Earth™ image below, we have combined the AD prevalence estimates with a KML file of counties in China sourced from GDAM (<http://gadm.org/>). The resulting output shows the coloured polygons of the counties in association with their projected AD prevalence as vertical data columns. This obviously contextualises the data and China's physical geography more effectively than the data columns alone. In addition we turned off the borders and labels for other countries in Google Earth so that the user can focus more directly on China and its data in preference to supporting contextual geographic data. This further illustrates the importance of careful methods selection and use in communicating complex health information, especially so when using a spatial health informatics environment. Too much information, if not absolutely essential, can be as confusing as too little.



Figure 5: Dual Layer Map of Counties and AD Prevalence Estimates for 2010

Finally, it is important to note that as more health systems develop and utilise 'big data' strategies, new tools and methods will be necessary to inform non-health audiences of the findings of such studies. In this context data visualisation is a relatively new but rapidly expanding field. Spatial data visualisation, as illustrated in these examples, is a key aspect of these emerging changes in the health informatics environment. The drive towards strategies such as tele-health and m-health, including methods such as the remote-monitoring of patients, will add to the necessity of having spatially-informed data that can be used to lobby for funding, resources, workforce and so forth.

Discussion

This project explores the likely impact of dementia and dementia sub-types in the People's Republic of China which is now the largest ageing society in the world. China is already utilising spatial science and technology to address age-related problems from a planning and service delivery perspective (e.g. Cheng et al, 2012) [5]. In addition, Yang et al [6] state that health and medical geography research is growing rapidly in China, indicating that ageing research and spatial methods will continue to converge. Lastly, several studies, such as the China Health and Retirement Longitudinal Study (CHARLS), are researching the prevalence and progression of the dementias in China [7]. This project adds to that complex emerging picture by modeling and visualising, in a spatial form, the scale of the issues that rapid population ageing presents for China and where its effects will be felt most strongly.

The value in this first iteration of our research is to make the totality of population ageing and its clinical and social effects accessible to a wide audience including various levels of government, policy developers, service planners, service providers and the broader citizenry. The need for societal level responses to the complexities and challenges presented by population ageing need to be supported by new and innovative instruments for monitoring and understanding the implications of current and future interventions. One such area of support lies in the new geographic and allied data visualisation tools becoming available to governments, planners, service providers and insurers.

The scale and speed of population ageing in China is profound and of an order of magnitude unlikely to be seen anywhere else except perhaps India. The multi-dimensional capacities of spatial science and technology represent a key domain for supporting societal-level responses to population ageing that aim for humane and sustainable outcomes. Ageing is fundamentally linked to space and place including the

connections people have to spaces and places as they age [7]. The future of population ageing requires a spatially informed and enabled approach to health information if such outcomes are to be achieved.

While the data we present here is shown at the county level, other administrative geographies can be utilised, such as province or prefecture. This means that, within the one information environment, data can be visualised for multiple scales and at different administrative levels. In addition, maps with more than one variable can be produced to example factors such as demography versus healthcare expenditure data, utilization of acute care services or the rates of other health conditions or events (cardiac, stroke, cancer etc.). Change is not only a feature of population ageing; it is a central feature of these spatial technologies. Their potential contribution to managing the complexity of health systems and health information management is on an upward developmental pathway. Which software to use and for what purposes will be highly contingent on the rate of change in this sector but the potential utility of such approaches remains clear and we can expect to see spatial methods added to other data visualisation environments.

Limitations

This is a point-in-time estimation of dementia prevalence and has the limitations of any similar modeling exercise. The aim of this model is not one of absolute accuracy, and we make no claims to that, but rather the application of spatial analysis and visualisation to a demographic trend with major implications for the Chinese government, healthcare systems, economy, and society. We have used the year 2010 as our reference year because of the production of the official Chinese census for that year, and recent revisions of dementia prevalence in China produced for the same year [8]. This makes the model reasonably accurate in a county-level approximation under those conditions, and a starting point for further investigation.

It is important to note, as indicated earlier, that research on dementia and their population-level effects is ongoing in China. Consequently, between this present work and the 2020 Chinese Census, we can expect to see a growing research information base come on stream. Simply put, because demography is so dynamic in China and researchers are working rapidly to build their understanding of population change and its effects, new data will almost certainly alter this picture and do so very soon. Its value lies in the visual dimension and in the technology's flexibility for exploring complex phenomena at the national level and below (provinces, counties etc.), all within the one virtual environment.

Conclusion

This developmental paper supports the growing awareness of the challenges that population ageing presents to China's health and social policy domains, as well as its rising number of elderly. The increasing survival of older people is a key achievement of China's health and economic developments of the past few decades, but systemic change will be necessary to meet the challenges posed by these successes. One key concern will be addressing the geographic scope of the inevitable challenges presented by population ageing and the rise of dementia across the country. These will co-occur with many other chronic and acute health conditions, making for new levels of complexity in identifying and treating the health needs of an ageing population.

We have presented a spatial virtual earth approach that makes a link between population ageing and dementia, which supports planning and response activities, as well as research

and interventional programs. While this is a pilot project, it seems clear that the complexities of China's physical geography, its population dynamics, and governance structures (central, provincial, city, local etc.) would benefit from a spatial strategy in order to cope with health and social care implications of ageing and dementia. In addition, we suggest that a spatially informed health informatics strategy can support policy design and applied responses to the challenges presented by population ageing. The future of health information management and population ageing responses are, we suggest, highly spatial, if current and emerging challenges are to effectively be addressed.

Acknowledgements

There are no acknowledgements or disclosures for this research.

References

- [1] United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. Working Paper No. ESA/P/WP.241. United Nations. New York. 2015
- [2] H. Guo, China's Earth observing satellites for building a Digital Earth. *International Journal of Digital Earth* 5(3) (2008), 185-188.
- [3] K.Y. Chan, W. Wang, J.J. Wu, L. Liu, E. Theodoratou, J. Car, L. Middleton, T.C. Russ, I.J. Deary, H. Campbell, W. Wang, and I. Rudan, Global Health Epidemiology Reference Group (GHERG). Epidemiology of Alzheimer's disease and other forms of dementia in China, 1990-2010: a systematic review and analysis. *Lancet* 381(9882) (2013), 2016-23.
- [4] Z. Liang and Q.J. Song, "Migration in China", In M.J. White (ed.), *International Handbook of Migration and Population Distribution*. Springer Handbook Series, Dordrecht, 2016.
- [5] Y. Cheng, J. Wang, and M.W. Rosenberg, Spatial access to residential care resources in Beijing, China. *International Journal of Health Geographics* 11 (2012), 32.
- [6] L. Yang, H. Li, Y. Cheng, and B. Wei, Medical and Health Geography. *In The Geographical Sciences during 1986 — 2015*. Springer, Singapore, 2017.
- [7] A. Davies and A. James, Geographies of Ageing: Social Processes and the Spatial Unevenness of Population Ageing. Ashgate, United Kingdom, 2011.
- [8] J.J. Pei, M.S. Giron, J. Jia, and H.X. Wang HX, Dementia studies in Chinese populations. *Neurosci Bull* 30(2), 2014, 207-16.

Address for correspondence

Hamish L. Robertson
robertsonhl@optusnet.com.au

Comparison of Regression Analysis and Transfer Function in Estimating the Parameters of Central Pulse Waves from Brachial Pulse Wave

Chai Rui^a, Xu Li-sheng^a, Yao Yang^a, Hao Li-ling^a, Qi Lin^a

^aSino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, China

Abstract

This study analyzed ascending branch slope (A_{slope}), dicrotic notch height (H_n), diastolic area (A_d) and systolic area (A_s), diastolic blood pressure (DBP), systolic blood pressure (SBP), pulse pressure (PP), subendocardial viability ratio (SEVR), waveform parameter (k), stroke volume (SV), cardiac output (CO), and peripheral resistance (RS) of central pulse wave invasively and non-invasively measured. Invasively measured parameters were compared with parameters measured from brachial pulse waves by regression model and transfer function model. Accuracy of parameters estimated by regression and transfer function model, was compared too. Findings showed that k value, central pulse wave and brachial pulse wave parameters invasively measured, correlated positively. Regression model parameters including A_{slope} , DBP, SEVR, and transfer function model parameters had good consistency with parameters invasively measured. They had same effect of consistency. SBP, PP, SV, and CO could be calculated through the regression model, but their accuracies were worse than that of transfer function model.

Keywords:

Pulse wave analysis, Regression analysis, Data accuracy

Introduction

Pulse wave contains abundant physiological and pathological information, which can reflect early changes of cardiac function parameters [1-2]. Information extracted from pulse wave can be used for clinical detection and treatment. A large number of studies have indicated that central artery is more accurate than brachial artery in predicting cardiovascular events [3]. However, since central pulse wave study is invasive, the direct measurement of its parameters is quite limited clinically. As far as the peripheral is concerned, brachial artery reflects cardiovascular events more accurately than radial artery does [4]. Therefore, the parameters of brachial pulse wave often indirectly reflect cardiovascular events. In addition to DBP and SBP, PP, k , SV, SEVR, CO, and R_s are also important indicators of cardiovascular function. Their meanings and calculation formulas are shown in Table 1. From the shape of the waveform, characteristic value of A_{slope} , H_n , A_s , A_d can also reflect the changes of cardiovascular function, and their descriptions [5-6] are shown in Figure 1. The ascending branch slope is defined as [7]:

$A_{\text{slope}} = (\text{Peak amplitude} - \text{Trough amplitude}) / (\text{Time corresponding to peak} - \text{Time corresponding to trough})$

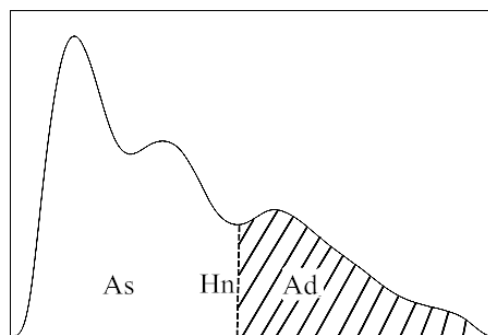


Figure 1- Characteristic value of pulse pressure waveform

Borow and Newburger [8] found that SBP and DBP of brachial artery and aorta were significantly related with a correlation coefficient equaling to 0.984 and 0.969. A linear correlation between brachial and aortic BP was reported, and it was used for the estimation of aortic BP by applying linear regression equations [8]. Then, similar significant correlation between aortic and brachial BP was proposed by Band et al [9]. However, in recent years, research on the regression relation of parameters is rarely reported. This paper presents the regression relationship between parameters of central pulse wave and brachial pulse wave, and makes a comparison with the transfer function model.

The generalized transfer function (GTF) model is widely used to estimate the parameters of central pulse wave [10, 11]. However, the regression model has the following advantages over the transfer function model: simple structures, few parameters, less sample requirements, and strong capacity of generalization that shows adequate performance of popularization to the new samples.

This paper analyzed some parameters that can reflect cardiovascular events effectively. The accuracy of the parameters, which were obtained by the regression model and transfer function model, was compared to explore whether the parameters of brachial pulse wave can effectively reflect the corresponding parameters of central pulse wave by establishing a linear regression model. This study not only plays a role of clinical diagnosis and treatment in the prediction of cardiovascular events, but also helps to develop new medical devices.

Table 1 - The physiological meaning and calculation formulas of parameters

PARAM	Function	Calculation formulas	Complementary
PP	PP can reflect central arterial stiffness and vascular aging more directly than other BP index [12].	$PP = SBP - DBP$	—
k	k value is an important index, which can predict cardiovascular disease continuously and independently [13].	$K = \frac{p_m - p_d}{p_s - p_d}$	$p_m = \frac{1}{T} \int_0^T p(t)dt$
SV	SV can send an early warning signal for cardiovascular disease [14].	$SV = \frac{0.283}{K^2} T(p_s - p_d)$	—
SEVR	SEVR is an index which can measure the balance between supply and demand of coronary arteries [15].	$SEVR = DPTI / TTI * 100\%$	DPTI (diastolic pressure-time index), TTI (tension-time index)
CO	CO is a basic indicator which can reflect the function of circulatory system [16][17].	$CO = \frac{17}{K^2} (P_s - P_d)$	—
R _s	The change of R _s is one of the important indexes of cardiovascular function[18].	$R_s = \frac{P_m}{CO}$	—

Materials and methods

Research materials

Patients who underwent coronary angiography in the First Hospital of China Medical University were selected for the study. A total of 40 patients were included in the study, 21 males and 19 females, age (54.76 ± 13.8 years), and weight (67.53 ± 8.5 kg). Invasive central pulse wave and brachial pulse wave were collected in interventional catheterization room. The pulse pressure waveforms of the aorta and brachial artery were collected synchronously with the pressure guide wires. Invalid waveform was excluded, and each volunteer provided a set of data for the study.

Instruments and data acquisition

Experimental instrument: St. Jude Medical Company's RadiAnalyzer Xpress, with sampling frequency of 100Hz.

Data acquisition procedures: In this study, the pressure waveforms of the ascending aorta and brachial artery were invasively measured by arterial catheter. The right radial artery of subjects was punctured using Seldinger technique in the supine position. Two pressure guide wires were placed about 2 to 3 cm above the elbow to calibrate the pressure. After calibration, one of the pressure guide wires was placed into the ascending aortic root, and the position of the other wire was unchanged. The invasive central pulse wave and brachial pulse wave were recorded when the pressure curve tended to be stable. The relevant parameters were calculated according to the pressure waveform. The relevant parameters were calculated by continuous five high quality pulse waveforms. The average of the five values was taken as the individual parameter.

Statistical and research methods

Parameters were extracted and calculated with MATLAB. Data were analyzed with SPSS software and Bland-Altman diagram was drawn with MedCalc software. Measurement data were expressed in " $\bar{x} \pm s$ ". Pearson correlation method was used to analyze the correlation. Paired t test and Bland-Altman analysis were used to test differences and consistency respectively.

A linear regression model was obtained with brachial pulse wave parameters (pA) and central pulse wave parameters (cA – the

"gold standard" [19]). Transfer function model was established with the waveform of the brachial pulse wave and the central pulse wave. New central pulse wave parameters (cA, gA) were obtained with the two models. The differences between cA and eA, cA and gA were tested with paired t test. The consistency between cA and eA, cA and gA was analyzed with Bland-Altman diagram. Finally, the results of the two models were compared. A $p < 0.05$ indicates that the difference was statistically significant.

Results

Correlation analysis and regression equations

The parameters were extracted from central pulse wave and brachial pulse wave according to the formula of the parameters. The correlation of each parameter was analyzed with Pearson correlation analysis method. The results showed that k value of central pulse wave and brachial pulse wave were not correlated ($p > 0.05$, $r < 0.3$), R_s were moderately correlated ($p < 0.01$, $r > 0.5$), and other parameters were highly positively correlated ($p < 0.01$, $r > 0.8$).

The regression equations were established according to correlation in the SPSS software, and the following equations were significant according to the criteria of regression analysis (parameters of central pulse wave are expressed as Y' and the corresponding parameters of brachial pulse wave are expressed as x). The regression equations are shown in Table 2.

Table 2 - The regression equation

PARAM	Regression equation
A_slope	$Y' = 0.509x + 1.453$
Hn	$Y' = 1.168x + 2.930$
As	$Y' = 0.934x + 509.954$
Ad	$Y' = 0.901x + 40.564$
DBP	$Y' = 0.932x + 3.846$
SBP	$Y' = 0.980x - 6.070$
PP	$Y' = 1.062x - 11.974$
SEVR	$Y' = 0.946x + 21.173$
SV	$Y' = 0.825x - 5.663$
CO	$Y' = 0.800x - 280.735$
R _s	$Y' = 1.515x - 0.820$

Parameter calculation results

The gold standard was the feature parameters extracted from the invasive central pulse wave. Regression model parameters (eA) are a new set of parameters of central pulse wave, which were calculated by linear regression models and parameters of brachial pulse wave (pA). Transfer function model parameters (gA) were a new set of central arterial parameters, which were calculated by the transfer function model (the establishment of a transfer function model refers to the contribution of Chen et al. [10] and Pauca et al. [11]).

Difference analysis with paired t test

The parameters eA and cA, gA and cA were tested by paired t test (gA and cA have been proved highly correlated ($p < 0.01$)), and the results are shown in Table 3. The results showed that there was no significant difference between eA and cA (A_slope, DBP, SBP, PP, SEVR, SV, CO, $p > 0.05$), so the above parameters could be obtained by regression model. There was no significant difference in all parameters of gA and cA.

Table 3 - Paired t test analysis of cA & eA and cA & gA

PARAM	cA & eA		cA & gA	
	t	p	t	p
A_slope	-.006	.995	.734	.468
Hn	-.006	.000*	-.042	.985
As	.012	.000*	-4.175	.937
Ad	-.002	.000*	3.152	.962
DBP	.124	.902	-.096	.924
SBP	-.006	.995	.009	.993
PP	-.006	.995	-.042	.966
SEVR	.022	.983	.003	.997
k	—	—	.004	.997
SV	-.557	.580	-.007	.995
CO	-.018	.985	.030	.976
Rs	10.269	.000*	0.17	.987

Note: * represents a significant difference

Consistency analysis with Bland-Altman analysis

The above two groups of parameters – cA and eA, cA and gA – do not have statistical difference. The two groups were analyzed by Bland-Altman analysis method. The scatter points of each parameter were in ($\bar{x} \pm 1.96SD$) range. The values of 95% limit of agreement are shown in Table 4. The Bland-Altman diagrams of A_slope, DBP and SEVR are shown in Figure 2, Figure 3, and Figure 4. In addition, cA and gA (A refers to Hn, Ad, As, k and Rs) also were analyzed. The Bland-Altman diagram of the five parameters showed that there were more than 95% scattered points distributed in range of $\bar{x} \pm 1.96SD$.

Table 4 - The value of 95% limit of agreement

PARAM	cA&eA	cA&gA
A_slope	(-1.61, 1.61)	(-1.57, 1.77)
DBP	(-3.8, 3.9)	(-3.9, 3.8)

SBP	(-9.3, 9.3)	(-5.6, 5.6)
PP	(-8.9, 8.9)	(-5.4, 5.4)
SEVR	(-13.5, 13.6)	(-13.6, 13.6)
SV	(-29.8, 32.6)	(-16.5, 16.5)
CO	(-2473.0, 2465.7)	(-885.1, 889.4)

The results showed that the parameters of central pulse wave (A_slope, DBP, SBP, PP, SEVR, SV, CO) calculated by the regression model had good consistency with the gold standard. Similarly, the parameters of transfer function model had good consistency with the gold standard. The consistency of the parameters (SBP, PP, CO, SV) calculated by the transfer function model was better than that of the regression model. Both the regression model parameters, including A_slope, DBP, SEVR, and the transfer function model parameters had good consistency with the parameters invasively measured, and their consistency had the same effect. The Hn, Ad, As, k value and Rs calculated by the transfer function model had good consistency with the gold standard.

Discussion and Conclusions

In this study, the linear relationship of the parameters between brachial pulse wave and central pulse wave was discussed, and the regression model was established through the waveform of the invasive central artery and brachial artery. The transfer function model was established as a contrast of the regression model. The eA and gA were compared to determine the accuracy of the regression model. To discuss whether reliable parameters of central pulse wave can be obtained by regression models, the following conclusions were obtained:

1. Clinical parameters of central pulse wave and brachial pulse wave (A_slope, Hn, As, Ad, DBP, SBP, PP, SEVR, SV, CO) were highly correlated, R_s was moderately correlated and k values were not correlated.
2. eA and cA (A_slope, DBP, SBP, PP, SEVR, SV, CO) did not have statistical difference ($p > 0.05$), and the consistency was good. Therefore, the above parameters could be obtained by regression model. However, Hn, As, Ad and R_s could not be estimated by regression model because of their statistical differences.
3. The consistency of A_slope, DBP and SEVR calculated by the two models was about the same, that was, eA could achieve the same effect as gA, and their regression equations were $Y' = 0.509x + 1.453$, $Y' = 0.932x + 3.846$, $Y' = 0.946x + 21.173$.
4. The consistency of gA (SBP, PP, CO, SV) and cA was better than eA and cA. However, whether the accuracy of the regression model will be improved needs to be further studied, in the case of a large sample.
5. Hn, As, Ad could not be estimated by regression model. From the formation mechanism of pulse wave, dicrotic notch is caused by the overlay of reflected wave caused by impact on peripheral arterial by ventriculus sinister ejection, and Dicrotic wave caused by flow of artery blood from telecentric end to proximal end, what is affected by factors like elasticity of vascular wall [20]. Another reason is that the influencing factors of blood vessels, such as vascular elasticity, have great individual differences. Therefore, Hn had a significant statistical differences and it could not be obtained by regression model. Similarly, the calculation of As and Ad were related to Hn.
6. k value is only affected by the shape of the pulse wave, which is irrelevant to blood pressure [14]. It will lead to the fact that two different waveforms probably have the same k value, which will be mistaken for same

physiological conditions of the two pulse waves. Therefore, k value needs to be directly calculated by the waveform, which could not be directly derived from the regression equation.

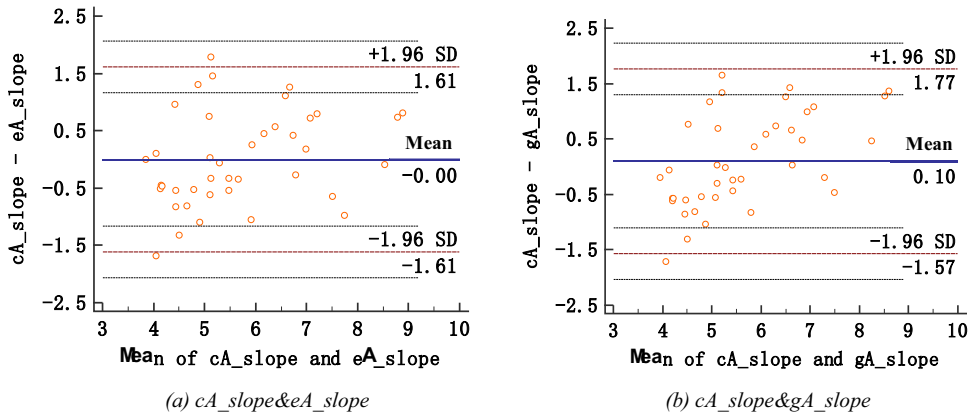


Figure 2 - The Bland-Altman analysis and comparison diagram of ascending branch slope

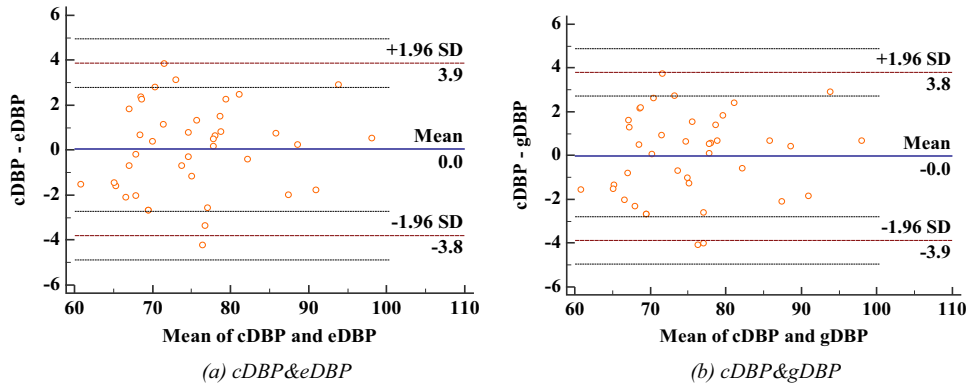


Figure 3 - The Bland-Altman analysis and comparison diagram of DBP

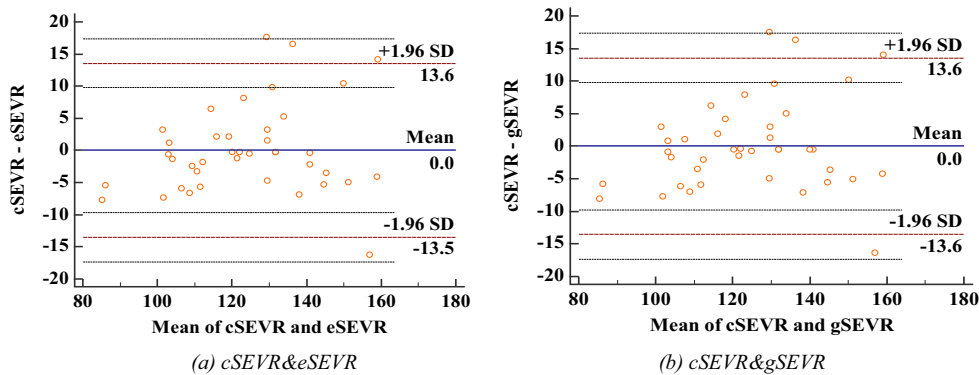


Figure 4 - The Bland-Altman analysis and comparison diagram of SEVR

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61374015, No. 61202258), the Ph.D. Programs Foundation of Ministry of Education of China (No. 20110042120037), the Liaoning Provincial Natural Science

Foundation of China (No.201102067) and the Fundamental Research Funds for the Central Universities (No. N110219001).

References

- [1] Y. Liu, Y. Yin, D. Zhang M. Wang, and Y. Bi, Simulation analysis of the pulse signal on the electricity network of cardiovascular system, *J Biomed Eng* 32(6) (2015), 1207-1211.

- [2] T. Pereira, C. Correia, and J. Cardoso, Novel methods for pulse wave velocity measurement, *J Med Biol Eng* **35**(5) (2015), 555-565.
- [3] M.J. Roeman, R.B. Devereux, J.R. Kizer, E.T. Lee, J.M. Galloway, T. Ali, J.G. Umans, and B.V. Howard, Central pressure more strongly relates to vascular disease and outcome than does brachial pressure the strong heart study, *Hypertension* **50**(1) (2007), 197-203.
- [4] W.Y. Li, X.H. Wang, L.C. Lu, Li, and H. Li, Discrepancy of blood pressure between the brachial artery and radial artery, *World J Emerg Med* **4**(4) (2013), 294-297.
- [5] L.S. Xu, M.Q.H. Meng, K. Wang, W. Lu and N. Li, Pulse images recognition using fuzzy neural network, *Expert Syst Appl* **36**(2) (2009), 3805-3811.
- [6] L.S. Xu, M.Q.H. Meng, C. Shi, K. Wang, and N. Li, Quantitative analyses of pulse images in Traditional Chinese Medicine, *Medical Acupuncture* **20**(3) (2008), 175-189.
- [7] J. Wang, X.L. Tian, X. Zhang, and M. Wang, Effects of fastigial nucleus stimulation on crucial cardiovascular physiological parameters, *Chin J Appl Physiol* **26**(4) (2010), 507-509.
- [8] K.M. Borow, J.W. Newburger, Noninvasive estimation of central aortic pressure using the oscillometric method for analyzing systemic artery pulsatile blood flow: Comparative study of indirect systolic, diastolic, and mean brachial artery pressure with simultaneous direct ascending aortic pressure measurements, *Am Heart J* **103**(5) (1982), 879-886.
- [9] J.I. Davies, M.M. Band, S. Pringle, S. Ogston and A.D. Struthers, Peripheral blood pressure measurement is as good as applanation tonometry at predicting ascending aortic blood pressure, *J Hypertens* **21**(3) (2003), 571-576.
- [10] C.H. Chen, E. Nevo, B. Fetis, P.H. Pak, F.C.P. Yin, W.L. Mughan, and D.A. Kass, Estimation of central aortic pressure waveform by mathematical transformation of radial tonometry pressure: Validation of generalized transfer function, *Circulation* **95**(7) (1997), 1827-1836.
- [11] A.L. Pauca, M.F. O'Rourke, and N.D. Kon, Prospective evaluation of a method for estimating ascending aortic pressure from the radial artery pressure waveform *Hypertension* **38**(4) (2001), 932-937.
- [12] S.S. Franklin, S.A. Khan, N.D. Wong, M.G. Larson, and D. Levy, Is pulse pressure useful in predicting risk for coronary heart Disease? The Framingham heart study *Circulation* **100**(4) (1999), 354-360.
- [13] T. Weber, M.F. O'Rourke, E. Lassing, M. Porodko, M. Ammer, M. Rammer, and B. Eber, Pulse waveform characteristics predict cardiovascular events and mortality in patients undergoing coronary angiography, *J Hypertens* **8**(4) (2010), 797-805.
- [14] L. Zhichang, Pulse contour method for the determination of cardiac output and its application in clinic practice, *Journal of Beijing Polytechnic University*, 1988.
- [15] D. Tsiachris, C. Tsioufis, D. Syrseloudis, D. Roussos, I. Tatsis, K. Dimitriadis, K. Toutouzas, E. Tsiamis, and C. Stefanadis, Subendocardial viability ratio as an index of impaired coronary flow reserve in hypertensives without significant coronary artery stenosis, *J Hum Hypertens* **26**(1) (2012), 64-70.
- [16] N. Mansencal, J. Delobelle, P. Balagny, M. Ihaddaden, M. Arslan, and O. Dubourg, Usefulness of a noninvasive cardiac output measurement using pulse wave transit time in coronary care unit, *Int J Cardiol* **168**(4) (2013), 4411-4412.
- [17] N. Fazeli, and J.O. Hahn, Estimation of cardiac output and peripheral resistance using square-wave-approximated aortic flow signal, *Frontiers in Physiology* **3** (2012), 298.
- [18] R.L. Hurt, Peripheral resistance, *Anesthesiology* **24**(2) (1963), 262.
- [19] R.M.L. Orme, D.W. Pigott, and F.G. Mihm, Measurement of cardiac output by transpulmonary arterial thermodilution using a long radial artery catheter. A comparison with intermittent pulmonary artery thermodilution, *Anaesthesia* **59**(6) (2004), 590-594.
- [20] W. Ping, D. Shuxian, and L. Yu-qing, The basic principle of Pulse Wave Analysis and its clinical applications, *Molecular Cardiology of China* **9**(6) (2009), 378-380.

Longitudinal Changes in Risk Stratification for a Managed Population

Lincoln Sheets, MD^{a,b}, Mihail Popescu, PhD^{a,b}, Kayson Lyttle, BSc^b,
Soo-Yeon Cho, PhD^b, Jerry C. Parker, PhD^b

^a Informatics Institute, University of Missouri, Columbia, Missouri, USA

^b School of Medicine, University of Missouri, Columbia, Missouri, USA

Abstract

The LIGHT² project managed the care of approximately 10,000 Medicare (primarily elderly) and Medicaid (low income) patients between 2013 and 2015. Risk tiers based on chronic disease diagnoses and recent healthcare utilization were strongly predictive of future healthcare utilization, and the authors expected that the members of an aging and well-insured population would gradually rise in risk of healthcare utilization over the course of three years. Various analytic techniques were used to characterize the members of higher risk tiers. However, retrospective cohort analysis and simple data visualization discovered the tendency of patients in lower initial risk tiers to remain healthy, and the tendency of patients in higher initial risk tiers to improve. In a time frame of three years, this return to stability was a more important influence on healthcare utilization than risk or aging.

Keywords:

Risk Assessment; Managed Care Programs; Data Display

Introduction

Managed Care and Risk Stratification

LIGHT² (Leveraging Information Technology to Guide Hi-Tech and Hi-Touch Care) was a Health Care Innovation Award from the Centers for Medicare and Medicaid Services to examine the use of advanced health information technology and care coordination in a managed population. With over 10,000 patient cases to be managed by fewer than 25 advanced practice nurses, the LIGHT² project needed clinical decision support to help focus preventive and longitudinal care on those patients who were at highest risk [1].

Definition and Validation of Risk Tiers

The LIGHT² investigators defined four risk tiers as (1) Healthy, (2) Stable, (3) Unstable, and (4) Complex, on the basis of diagnoses that are included in the Chronic Conditions Data Warehouse (CCW) [2] and on healthcare utilization in the preceding year (see Table 1). These risk tiers, based on simple retrospective data, were useful predictors of the prospective risk of patient healthcare utilization [3]. The authors expected that the members of an aging and well-insured population would gradually rise in risk of healthcare utilization during the course of the study period.

Objective

The objective was to demonstrate and visualize the longitudinal changes in risk stratification within a population of well-insured primary care patients.

Methods

LIGHT² Population and Setting

The University of Missouri Health System (UMHS) enrolled over 10,000 primary-care patients with Medicare or Medicaid coverage in the LIGHT² study between early 2013 and early 2015.

Data Sources

Diagnoses and utilization histories for all LIGHT² patients between 1 January 2012 and 31 December 2014 were collected retrospectively from the UMHS Health Analytics Library, an analytic data mart created for the LIGHT² study to reflect the relevant contents of the UMHS electronic health records system.

Visualization of Movement between Risk Tiers

In order to visualize the longitudinal stability of the four risk tiers in the study population, cohorts were retrospectively identified by the risk tier of each patient at the beginning of the study period. The risk tiers were then retrospectively recalculated bi-weekly for each patient in the cohort over the following three years. The percentage of each cohort in each tier, as well as the percentages deceased or lost to follow up, were calculated and displayed as stacked bar charts.

Results

“Healthy” (Tier 1) patients had a large chance (56%) of staying in the same risk tier, and most of remainder (27%) moved only to the next higher tier (see Figure 1). “Stable” (Tier 2) patients had a very large chance (68%) of staying in the same risk tier, and half of remainder (13%) moved only to the next higher tier (see Figure 2). “Unstable” (Tier 3) patients had a large chance (48%) of moving to the next lower tier, and most of remainder (26%) stayed in the same risk tier (see Figure 3). “Complex” (Tier 4) patients had a fairly large chance (37%) of moving to a much lower risk tier, and almost half of remainder (25%) moved to the next lower tier, leaving a fairly small chance (18%) of remaining in the same tier (see Figure 4).

The rate of death increased steadily in higher risk tiers, at 1%, 4%, 7%, and 9% respectively (see Figures 1-4). The rate of loss to follow up was fairly consistent across tiers, at 12%, 9%, 10%, and 11% respectively (see Figures 1-4).

Discussion

Contrary to the authors' hypothesis, patients in lower risk tiers ("Healthy" and "Stable" patients) at the beginning of the study period were most likely to remain in lower risk tiers after three years, and those in higher initial risk tiers ("Unstable" and "Complex" patients) were most likely to move to lower risk tiers after three years.

The consistently higher death rate in higher risk tier cohorts provides an additional validation of the predictive utility of the risk tiers.

Because initial validation of the tiers showed that "Complex" (Tier 4) risk stratification was associated with significantly higher utilization [3], additional data science techniques were used to characterize the "Complex" patients, including contrast mining. For this analysis, patients were initially stratified by high cost and low cost, and those among the most expensive 5% were found to incur 50% of costs. Complex (Tier 4) patients were 4-5% of the total managed population at any given time, therefore roughly equivalent to the 5% high-cost patients. Association rule mining [4] was used to find all possible combinations of 19,014 sparse attributes. Those found frequently ($Pr \geq 0.2$) in 479 high-cost patients but infrequently ($Pr < 0.2$) in equal numbers of low-cost patients were tested for statistical significance, revealing eleven associations that characterized the high-cost patients (see Table 2) [5].

Although contrast mining found distinct patterns of utilization within the "Complex" (Tier 4) patients, less than one fifth of that cohort remained in Tier 4 by the end of the three-year study period, with another fifth deceased or lost to follow up. The remaining three fifths of the cohort had returned to the lower "Stable" (Tier 2) and "Unstable" (Tier 3) risk tiers.

Limitations

Because the data were collected from a single healthcare system, some utilization of other providers may be missing. This could affect both the risk stratification of patients and the number of their utilization episodes after stratification. This data gap could be resolved by supplementing electronic medical records with insurance claims data for the same population over the same period.

Future Research

This study measured the net number of patients in each risk tier at each point in time, but did not distinguish how many individuals move in each direction made up this net result. To gain a greater insight into these movements, hidden Markov models could be used to track the tiers or "states" through which individual patients moved. Contrast mining, which has been applied successfully to the "Complex" (Tier 4) sub-population, should also be applied to patients in all risk tiers. These methods and sequence mining techniques are needed to find the characteristics of patients in each tier that move to lower or higher tiers, as well as those who remain in their initial tiers.

Conclusion

Despite the use of several sophisticated analytic techniques, including deeper understanding of high-risk patients using

contrast mining, simple data visualization discovered a remarkable pattern in this study population: the tendency of patients at lower initial risk to remain healthy, and the tendency of patients at higher initial risk to improve. In the time frame of a few years, this return to stability was a more important influence on healthcare utilization than risk or aging. The expected gradual rise in risk tiers, which was not seen in these results, may be detectable with longer longitudinal data.

Authors may use tables to display data and results.

Table 1 – Definitions of Risk Tiers

Tier	Definition
1: Healthy	No CCW* diagnoses
2: Stable	One or more CCW* chronic conditions AND (hospital episodes = 0 AND outpatient visits < 5) in prior year
3: Unstable	One or more CCW* chronic conditions AND (hospital episodes = 1 OR outpatient visits = 5 to 12) in prior year
4: Complex	One or more CCW* chronic conditions AND (hospital episodes > 1 OR outpatient visits > 12) in prior year

Table 2 – Contrast-Mined Attributes Significantly ($p < 0.05$)
Associated with High-Cost Patients

Attribute	Odds Ratio
ICD9=311 depressive disorder	1.707
ICD9=401.9 unspecified essential hypertension	1.423
ICD9=414 ischemic heart disease	1.828
ICD9=715 osteoarthritis	2.769
Obesity (BMI > 30)	9.496
Anti-infectives	1.504
Benzodiazepines	1.307
Beta-adrenergic blocking agents	1.314
Quinolones	1.674
Respiratory agents	1.340
Selective serotonin reuptake inhibitors	0.655

* Intercept = -4.2585 with $p < 0.0001$

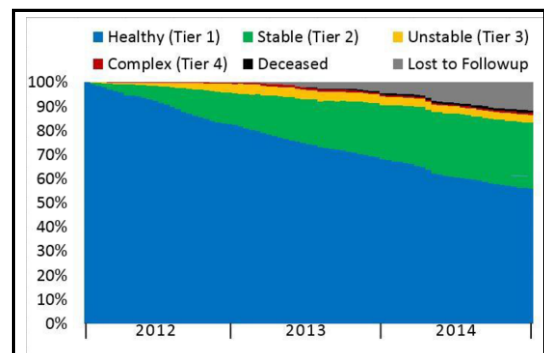


Figure 1 – Movement of the Healthy (Tier 1) Cohort in the Following Three Years

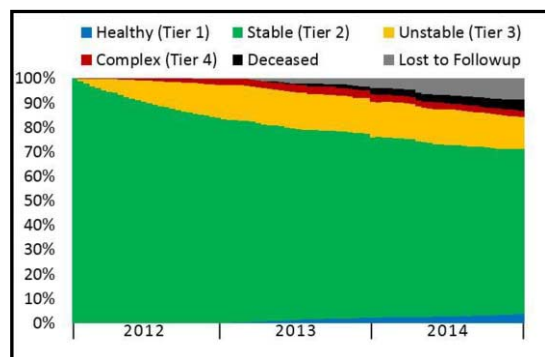


Figure 2 – Movement of the Stable (Tier 2) Cohort in the Following Three Years

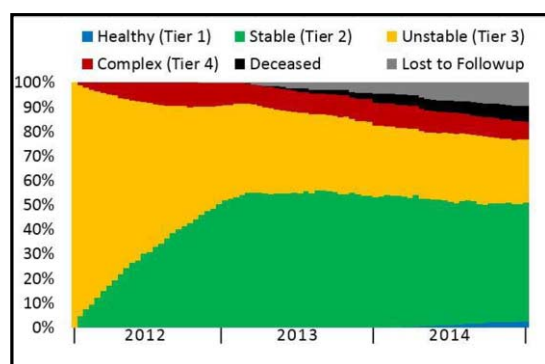


Figure 3 – Movement of the Unstable (Tier 3) Cohort in the Following Three Years

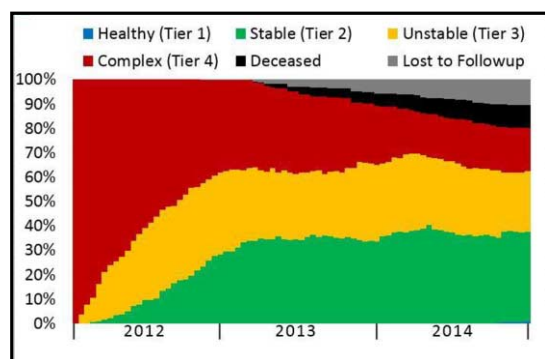


Figure 4 – Movement of the Complex (Tier 4) Cohort in the Following Three Years

Acknowledgements

This publication was made possible by Grant Number 1C1CMS331001-01-00 from the Department of Health and Human Services, Centers for Medicare & Medicaid Services. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services or any of its agencies. The funding agreement ensured the authors'

independence in designing the study, interpreting the data, writing, and publishing the report. The research presented here was conducted by the awardee. Findings might or might not be consistent with or confirmed by the independent evaluation contractor.

The authors would also like to thank these investigators for their contributions to this study:

Chi-Ren Shyu, PhD Greg Petroski, PhD Yan Zhuang, BS Bin Ge, MA, MD

References

- [1] L. L. Popejoy, J. Jaddoo, J. Sherman, C. Howk, R. Ngu-yen, and J. C. Parker, Monitoring Resource Utilization in a Health Care Coordination Program, *Professional Case Management*, **20** (2015), 310-320.
- [2] R. A. Goodman, S. F. Posner, E. S. Huang, A. K. Parekh, and H. K. Koh, Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice, *Preventing Chronic Disease*, **10** (2013) 1-16.
- [3] L. Sheets, L. L. Popejoy, M. Khalilia, G. F. Petroski, and J. C. Parker, Identification of Patients at Risk for High Utilization of Healthcare Services, *AMIA Annual Symposium Proceedings*, **2016** (2016), in press.
- [4] R. Agrawal and R. Srikant, Fast Algorithms for Mining Association Rules in Large Databases, *Proceedings of the 20th International Conference on Very Large Data Bases*, **20** (1994), 487-499.
- [5] L. Sheets, G. F. Petroski, M. A. Phinney, Y. Zhuang, B. Ge, J. C. Parker, and C. R. Shyu, Combining Contrast Mining with Logistic Regression to Predict Healthcare Utilization in a Managed Care Population, *Applied Clinical Informatics*, in press.

Address for correspondence

Lincoln Sheets, University of Missouri, Columbia, Missouri, USA
 telephone: 417-860-1197; fax: 573-884-4808
 email: SheetsLR@health.missouri.edu

Phenotypic Analysis of Clinical Narratives Using Human Phenotype Ontology

Feichen Shen, Liwei Wang, Hongfang Liu

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

Phenotypes are defined as observable characteristics and clinical traits of diseases and organisms. As connectors between medical experimental findings and clinical practices, phenotypes play vital roles in translational medicine. To facilitate the translation between genotype and phenotype, Human Phenotype Ontology (HPO) was developed as a semantically computable vocabulary to capture phenotypic abnormalities found in human diseases discovered through biomedical research. The use of HPO in annotating phenotypic information in clinical practice remains unexplored. In this study, we investigated the use of HPO to annotate phenotypic information in clinical domain by leveraging a corpus of 12.8 million clinical notes created from 2010 to 2015 for 729 thousand patients at Mayo Clinic Rochester campus and assessed the distribution information of HPO terms in the corpus. We also analyzed the distributional difference of HPO terms among demographic groups. We further demonstrated the potential application of the annotated corpus to support knowledge discovery in precision medicine through Wilson's Disease.

Keywords:

Semantic Annotation; Human Phenotype Ontology; Phenotypic Analysis

Introduction

Phenotypes, defined as observable characteristics and clinical traits of diseases and organisms, have attracted increasing attention in the area of translational medicine by serving as the connectors between medical experimental findings and clinical practices. For example, in genomic medicine, phenotypes provide evidence to stratify and differentiate various groups of patients in order to discover specific hidden genotype-phenotype associations [1].

As a tool for annotating human phenotypic abnormalities, Human Phenotype Ontology (HPO) [2] has been developed as a controlled vocabulary for phenotypes by mining and integrating phenotype knowledge from medical literatures, Orphanet [3], Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) [4], and Online Mendelian Inheritance in Man (OMIM) [5]. HPO also provides associations with other biomedical resources such as Gene Ontology [6]. Multiple studies have utilized HPO for data annotation. For example, Taboada [7] developed a semantic annotation system to automatically extract rare disease information from PubMed literature using HPO. In addition, Westbury and his colleagues [8] used a HPO-based clustering approach to identify group of heritable bleeding and platelet disorders (BPD) cases. Zhu and her colleagues [9] leveraged HPO to develop a genetic testing knowledge base.

Here, we used HPO to annotate a large collection of clinical narratives consisting of all clinical notes generated from 2010 to 2015 at Mayo Clinic Rochester campus and assessed the distributional information of HPO terms. A case study was performed to demonstrate the potential application of the annotated data.

In the following, we describe the methods used for generating the annotated corpus. Next, we present and discuss our results. We then analyze the distributional differences among demographic groups and detail the case study followed by the conclusion and potential future directions.

Background and Materials

Clinical Data Collection

We collected all clinical notes created between 2010 and 2015 at Mayo Clinic Rochester campus with research authorization. The resulting corpus contains 12.8 million clinical notes corresponding to 729 thousand patients. Specifically, we limited our annotation to the diagnosis section of clinical notes for better extracting phenotype related terms.

The Unified Medical Language System and MetaMap

The Unified Medical Language System (UMLS) [10] is an integrated database that contains all key medical terminologies and their related resources. Each term in the UMLS has a Concept Unique Identifier (CUI) and belongs to a specific semantic type. MetaMap [11] is a configurable application that is able to map biomedical terms to UMLS Metathesaurus. In this study, we applied MetaMap with the UMLS version 2015 on free-text clinical notes to extract related terms as a preprocessing step.

Human Phenotype Ontology

We used the latest version of Human Phenotype Ontology (HPO) released on September 2016. The HPO contains four sub-ontologies as shown in Table 1, and each sub-ontology has its different focus on annotating area.

HPO package contains an ontology file that records all HPO phenotype terms with their synonyms and resource references (e.g., the UMLS). As shown in Figure 1, in the ontology hierarchical structure, *Menorrhagia* (HP_0000132) belongs to sub-ontology *Phenotypic Abnormality* and is listed as a subclass of *abnormal bleeding*. Two synonyms *abnormally heavy bleeding during menstruation* and *hypermenorrhea* are connected to it through object property *has_exact_synonym*. Moreover, *Menorrhagia* relates to a UMLS CUI C0025323 through Web Ontology Language (OWL) axiom and object property *annotatedTarget*. The corresponding Resource Description Framework (RDF) representation in Figure 1 indi-

cates such triple relationships. Although HPO has link to the UMLS, the current version of HPO has not been fully incorporated into the UMLS version after 2013. Due to this limitation, the combination of both HPO and the UMLS is able to make a more comprehensive annotation than using only one of them.

In addition, there is another clinical annotation file wrapped in HPO package that captures around 119,000 phenotype-disease associations where a majority of them (115,000) correspond to hereditary conditions. Take *Wilson's Disease* as an example, the HPO annotation file records 33 relevant phenotypes, e.g., *Proteinuria*, *Dementia*, and *Coma* etc.

Table 1–HPO Sub-ontologies

Sub-ontology	Description
Phenotypic Abnormality	Main ontology of the HPO and covers most of the clinical abnormalities
Mode of Inheritance	Sub-ontology to describe mode of inheritance
Clinical Modifier	Sub-ontology with description of typical modifiers of clinical symptoms
Mortality/Aging	Sub-ontology that indicates time of death

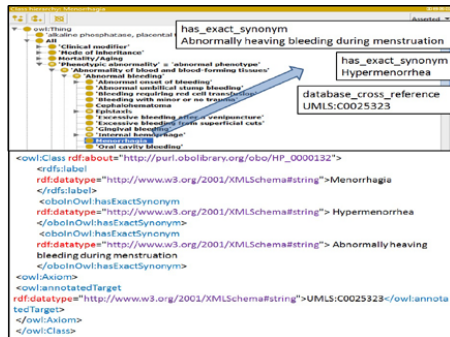


Figure 1– An Example of HPO Phenotype

Methods

Our annotation workflow is designed to leverage the UMLS and the corresponding NLP tool MetaMap. Figure 2 shows the overview of our annotation workflow, a combination of natural language processing (NLP) and semantic processing techniques. It includes two modules: i) the preprocessing module which leverages NLP techniques to extract key words from text; and ii) the semantic annotation module which maps the lexicon entries acquired from text to HPO terms and applies a greedy approach for each entry.

NLP Preprocessing Module

We first identify UMLS concepts in the diagnosis section of the clinical notes using MetaMap. We only keep UMLS concepts with the following phenotype-related semantic types: Disease or Syndrome (dsyn), Neoplastic Process (neop), Mental or Behavioral Dysfunction (mobd), Anatomical Abnormality (anab), Congenital Abnormality (cgab), Injury or Poisoning (inpo), Finding (findg) and Sign or Symptom (sosy).

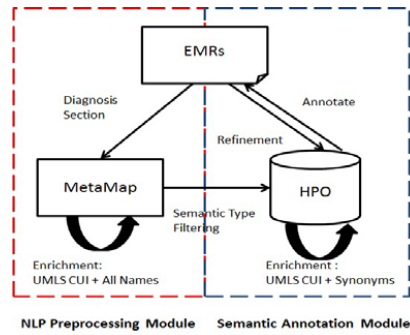


Figure 2– Annotation Work Flow

Semantic Annotation Module

Our annotation focuses on the main HPO sub-ontology *Phenotypic Abnormality* (HP_0000118) and its descendants. It contains 11,721 phenotype terms and 19,358 related synonyms defined in HPO or inferred from the linked UMLS CUIs. Our mapping includes two steps: we map UMLS concepts derived from the NLP preprocessing module to HPO terms using exact match, in case of ambiguity, we select the one with the smallest CUI number; and for HPO terms that failed to be mapped to the UMLS, the module then conducts another round of string matching refinement. Specifically, it reviews tokenized original text to check if there exists any possible candidate missed by MetaMap.

Results

The proposed annotation system was developed with Eclipse Standard/SDK version Luna 4.4.0. The interface to access MetaMap, the UMLS, and HPO were coded in Java programming language. In addition, we deployed the whole system on the Open Grid Scheduler (OGS) framework running on 64 bit Linux CentOS 6.8 servers hosted by Mayo Clinic to improve processing speed on large amounts of clinical notes.

In the following, to protect patients' privacy, any phenotypes with counts less than 11 were marked as <0.0015% or <11.

HPO Phenotype Coverage in Clinical Data

Overall, 3,241 out of 11,721 (27.7%) HPO phenotypes were extracted from 6,771,409 (52.8%) clinical notes. HPO has 23 abnormality categories. In our cohort, we counted the number of unique phenotypes and computed phenotype coverage for each category as shown in Figure 3. For example, *Abnormality of the nervous system* has the most phenotypes involved (599), e.g., *Anomia*, *Anxiety*, *Bone pain* and *Dysesthesia* while *Abnormality of the thoracic cavity* has no phenotype contained. Meanwhile, *Abnormality of the voice* holds a small group of phenotypes (11), including *Dysphonia*, *High pitched voice*, *Hoarse cry*, and *Laryngeal dystonia* etc. However, in terms of phenotype coverage, 78.6% (11 of 14) phenotypes under *Abnormality of the voice* were found. *Abnormality of the breast* also holds the second highest phenotype coverage as 55.6% (15 of 27). Since one phenotype can belong to more than one abnormality categories, in clinical notes, we found that 846 phenotypes are marked by 2 categories, 157 phenotypes dropped in 3 categories, 15 phenotypes are depicted by 4 categories, and only 1 phenotype is described by 5 categories.

Table 2 shows the top 10 frequent phenotypes as well as 10 phenotypes with relatively less frequent occurrences in clinical notes. We can see that many HPO common phenotype terms were found in cohort, such as *Hypertension*, *Hyperlipidemia*, *Apnea*, or *Anxiety*, etc.

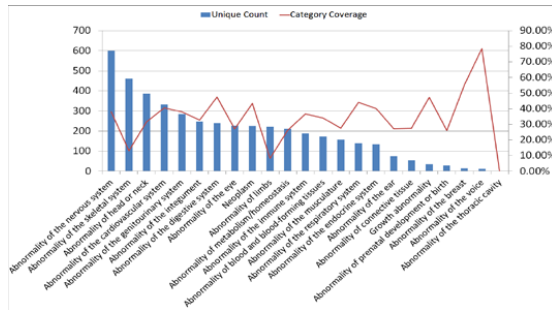


Figure 3–Distribution and Coverage for HPO Categories

We also noted our clinical notes contain phenotypes that related to rare diseases. For example, *Alacrima* is one of the features for *Triple A syndrome*. *Hemifacial hypertrophy* is a rare congenital disease. *Episodic tachypnea* often comes along with *Joubert syndrome*. *Alobar holoprosencephaly* is associated with *Holoprosencephaly*. *Barrel-shaped chest* is one of the symptoms of *Dyggve Melchior Clausen (DMC) syndrome*. *Chemodectoma* contributes to one of the signs caused by *Abdominal Chemodectomas with Cutaneous Angiolipomas*.

Table 2– Selected HPO Annotated Common/Rare Phenotypes

Common Phenotypes	Patients' Percentile	Rare Phenotypes	Patients' Percentile
Hypertension	19.8%	Alacrima	<0.0015%
Hyperlipidemia	15.2%	Intercostal retractions	<0.0015%
Depression	9.1%	Lattice retinal degeneration	<0.0015%
Apnea	8.6%	Hemifacial hypertrophy	<0.0015%
Back pain	8.5%	Decreased lacrimation	<0.0015%
Obesity	8.4%	Episodic tachypnea	<0.0015%
Sleep apnea	8.2%	Alobar holoprosencephaly	<0.0015%
Anxiety	8.2%	Barrel-shaped chest	<0.0015%
Headache	8%	Chemodectoma	<0.0015%
Arthritis	7.7%	Gastrointestinal carcinoma	<0.0015%

Demographic Phenotype Analysis

We retrieved patients' associated demographic information and assessed the phenotype distributional difference among demographic groups.

We analyzed the distribution of HPO abnormality categories among different age groups. Basically, we grouped age 0-17, 18-35, 36-54 and >55 as children and teenagers, young adults,

mid adults, and old adults age groups. As shown in Figure 4, *Abnormality of the nervous system* is popular in each age group. *Abnormality of immune system* occurs more frequently in children and teenagers. *Abnormality of the digestive system* often comes with young adult patients. While *Abnormality of the cardiovascular, skeletal, metabolism/homeostasis and respiratory systems* are common problems for mid adults but more common problems for old adults. In addition, *Immune system disorder*, *Abnormality of genitourinary system*, and *Neoplasm* have a high prevalence in old patients.

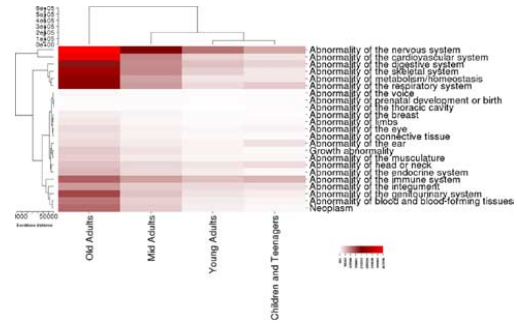


Figure 4–Distribution of Phenotypes across 4 Age Groups

In total, there are 2,864 unique phenotypes found in male and 2,907 unique phenotypes detected in female. Based on patients' percentile for both genders, we found that *Hypertension*, *Apnea*, *Hyperlipidemia*, *Back pain*, and *Obesity* are 5 most frequently occurred common phenotypes. We also conducted two case-control studies on male and female respectively to discover dominant phenotypes for each gender. We ranked phenotypes by descending order of odds ratios and filtered out phenotypes that are only associated with one specific gender (e.g., *Cryptorchidism* for male and *Amenorrhea* for female). With the odds ratio of 13.6, *Hypergonadotropic hypogonadism* is the phenotype that closely related to male (Table 3). Meanwhile, *Hirsutism* shows the highest odds ratio of 94.3 with female (Table 4). In general, top phenotypes in female held a relative higher average odds ratio.

Table 3– Top 5 Male-Specific Phenotypes

Phenotype	#Patients	odds ratio	p-value
Hypergonadotropic hypogonadism	136	13.6	2.0e-31
Violent behavior	20	11	3.9e-05
Dissecting aortic aneurysm	18	9.9	0.1e-03
Calf muscle hypertrophy	<11	9.9	0.9e-02
Sinusitis	9975	9.4	3.7e-158

Table 4– Top 5 Female-Specific Phenotypes

Phenotype	#Patients	odds ratio	p-value
Hirsutism	1140	94.3	6.4e-298
Hepatocellular adenoma	114	26	4.9e-27
Bulimia	417	23.7	3.4e-94
Choroid plexus cyst	126	16.4	4.6e-27
Neural tube defect	55	12.5	1.8e-11

We also extracted 29 races from clinical notes and picked 2 with the most population (excluding other and unknown races), which are White (519,098) and Black or Africa (8,835). Based on phenotype frequencies, we found top 5 common phenotypes for both races are *Hypertension*, *Apnea*, *Hyperlipidemia*, *Depression*, and *Back pain*, which highly overlapped with top 5 common phenotypes for male and female except *Depression*. We also conducted two case-control studies on White and Black or Africa patients to acquire significant phenotypes for each race. Table 5 shows top 5 significant White-specific phenotypes based on odds ratio. Based on literature review, we validated that *Basal cell carcinoma*, *Melanoma*, and *Macular degeneration* are highly related to patients in white race [12-14]. Table 6 shows top 5 significant Black or Africa-specific phenotypes based on odds ratio with the number of affected patients greater than or equal to 3. We found that 4 out of 5 top Black or Africa-specific phenotypes are related to genetic and rare diseases. *Scarring alopecia of scalp* is a rare disorder that destroys hair follicle. *Clitoral hypertrophy* is a congenital malformation and a rare condition. *Intellectual disability, profound* is related to *Lissencephaly*. *Spastic ataxia* is a symptom of *Autosomal Dominant Hereditary Ataxia*. Except those rare phenotypes, *Status asthmaticus* is a severe asthma. Literature indicated that *Scarring alopecia of scalp* and *Intellectual disability* are related to Black or Africa patients [15-16]. For other 3 phenotypes, even they occur in few patients but it may be worth to explore more as little evidence has been found about the relationship between them and Black or Africa patients in literature.

Table 5–Top 5 White-Specific Phenotypes

Phenotype	#Patients	odds ratio	p-value
Basal cell carcinoma	12969	113.2	3.2e-92
Melanoma	20235	44.8	5.6e-136
Barrett esophagus	2899	16.5	1.9e-17
Telangiectasia	1884	16.1	1.6e-11
Macular degeneration	1330	11.3	8.5e-08

Table 6–Top 5 Black or Africa-Specific Phenotypes

Phenotype	#Patients	odds ratio	p-value
Scarring alopecia of scalp	<11	25.2	0.5e-03
Clitoral hypertrophy	<11	14.7	0.3e-03
Intellectual disability, profound	<11	13.6	0.2e-02
Status asthmaticus	<11	11.8	0.1e-03
Spastic ataxia	<11	11.8	0.3e-02

A Case Study of Phenotypic Analysis

Wilson's Disease (WD) is an autosomal recessive rare inherited disorder that makes copper accumulate in organs, and its diagnosis can be challenging as its signs and symptoms are often indistinguishable from those of other liver diseases, such as hepatitis. In this section, we conducted two case-control studies to demonstrate the potential of HPO-annotated data in supporting disease diagnosis. Among a total of 615,590 patients, we extracted 39 patients with confirmed diagnosis of WD and 63 patients suspected for WD but negative by genetic testing, leaving 615,488 patients with no clinical assertions or doubts for WD. Based on the three groups, two case-control studies were designed as follows: 1) To reveal significant phenotypes of WD, odds ratios of phenotypes were computed for 39 confirmed WD patients compared to 615,488 Non-WD patients; 2) To investigate the phenotypes significant for similar disease differentiation, odds

ratios of phenotypes were computed for 39 confirmed WD patients compared to 63 patients suspected for WD.

Table 7 and 8 shows top 10 phenotypes ranked by odds ratio (with the number of affected patients greater than or equal to 3) representing significant clinical features of WD patients against the general population and the specific cohort with similar symptoms, respectively. Overall, odds ratio of phenotypes in the WD cohort against the general population were higher. Although the number of patients was relatively low, manual evaluation showed that all phenotypes are related to WD [17-19]. In contrast, odds ratio of phenotypes in the WD cohort against the similar cohort were relative lower, reflecting the reality that differential difficulties exist in the diagnosis of WD and WD similar diseases. However, according to [17], 7 phenotypes *Hypoalbuminemia*, *Nevus*, *Osteopenia*, *Cognitive impairment*, *Cirrhosis*, *Renal insufficiency*, and *Fever* are common symptoms of WD, of which only *Osteopenia* and *Cirrhosis* are covered by WD-related phenotypes recorded in HPO annotation file. The remaining are possible comorbidities highly occurring with WD. This analysis revealed the potential of these phenotypes in differentiating WD from similar diseases.

Table 7–Top 10 Phenotypes with Odds Ratio for Confirmed WD Patients to Non-WD related Patients

Phenotype	#Patients	odds ratio	p-value
Hepatic failure	<11	22.5	0.4e-03
Hypoalbuminemia	<11	12.1	0.2e-02
Cirrhosis	11	10.9	2.0e-08
Elevated hepatic trans-aminases	<11	7.29	0.9e-02
Splenomegaly	<11	7.27	0.9e-02
Ascites	<11	5.52	0.7e-02
Encephalopathy	<11	4.54	0.6e-02
Hepatitis	<11	4.02	0.5e-02
Lipoma	<11	3.51	0.6e-01
Thrombocytopenia	<11	3.18	0.9e-02

Table 8–Top 10 Phenotypes with Odds Ratio for Confirmed WD Patients to Suspected Patients

Phenotype	#Patients	odds ratio	p-value
Hyperglycemia	<11	5.2	0.2
Hypoalbuminemia	<11	5.2	0.2
Nevus	<11	3.5	0.2
Esophagitis	<11	2.5	0.4
Osteopenia	<11	2.5	0.4
Cognitive impairment	<11	2.5	0.4
Hyperlipidemia	<11	2.54	0.2
Cirrhosis	11	2.08	0.2
Renal insufficiency	<11	2.06	0.3
Fever	<11	1.71	0.5

Discussion

Our study showed that almost half of the clinical notes contain no HPO terms. To investigate further, we extracted top 10 frequent phenotypes from those notes including *Fatigue*, *Pharyngitis*, *Dyslipidemia*, *Dermatoheiosis*, *Insomnia*, *Diabetes*, *Fibromyalgia*, *Snoring*, *Chronic Obstructive Airway Disease*, and *Rash*. In general, most of these phenotypes are not hereditary ones, and some of them are more generic than HPO terms (e.g., SNOMED CT terms [20]). For example, *Fatigue*

covers HPO annotated phenotype *Exercise-induced muscle fatigue*, *Pharyngitis* gives a more general description of HPO annotated phenotype *Recurrent Pharyngitis*, and *Diabetes* is the super set of HPO annotated phenotype *Insulin-resistant diabetes mellitus*.

Note that in our case study, since *Wilson's Disease* is a rare disease, we were not able to provide a large sample size, which results in relative low odds ratio and high p-value when compared confirmed patients to the general population and suspected patients. In addition, in this research, to reduce the confounding factor of comorbidity, we picked phenotypes at the same year with WD diagnosis. In the future, to further address this issue, a more specific timestamp can be added to put restriction on diagnosed period. Currently, HPO records 33 phenotypes specifically related to WD. Although HPO coverage for WD is not sufficient, phenotypes in HPO can help on extracting all symptoms of diseases, making it possible to reveal significant phenotypes of WD by conducting case-control analysis. Therefore, it is possible to combine clinical notes with HPO to provide empirical evidences to further help diagnose and differentiate WD from similar diseases.

The integration of genotypic and phenotypic knowledge is an essential step to facilitate knowledge discovery in translational medicine. HPO, as a tool with linkage to gene databases and statistics of patients' percentile of phenotypes for each disease (e.g., 1% as very rare, 5% as rare), can be leveraged for this purpose. Especially for rare phenotypes, HPO makes it convenient for researchers to explore the potential genotype-phenotype associations.

Conclusions and Future Work

In this study, we used HPO to annotate phenotypes in clinical text for conducting phenotypic analysis. We demonstrated its potential in facilitating knowledge discovery.

In the future, we will try different ontologies and dictionaries to detect phenotypes from clinical notes and compare performances among them. Furthermore, a more specific diagnosis time range will be considered to largely reduce mischaracterization caused by complication and comorbidity. Moreover, we will use HPO with NLP and machine learning techniques to link various phenotypes with genes and drugs to perform knowledge discovery on gene disease and drug repositioning.

Acknowledgements

We would like to thank the support from our colleague Ravikumar Komandur Elayavilli for his valuable suggestions on this research. This work was made possible by internal funding from Center for Individualized Medicine of Mayo Clinic and NCATS Biomedical Translator Award, OT3TR0021901.

References

- [1] Denny, Joshua C., et al. PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26** (2010), 1205-1210.
- [2] Robinson, Peter N., et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics* **83** (2008), 610-615.
- [3] Weinreich, Steffanie S., et al. Orphanet: a European database for rare diseases. *Nederlands tijdschrift voor geneeskunde* **152** (2008), 518-519.
- [4] Firth, Helen V., et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *The American Journal of Human Genetics* **84** (2009), 524-533.
- [5] Hamosh, Ada, et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research* **33**, suppl 1 (2005): D514-D517.
- [6] Ashburner, Michael, et al. Gene ontology: Tool for the unification of biology. *Nature Genetics* **25** (2000), 25-29.
- [7] Taboada, Maria, et al. Automated semantic annotation of rare disease cases: a case study. *Database* **2014** (2014): bau045.
- [8] Westbury, Sarah K., et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. *Genome Medicine* **7** (2015), 1.
- [9] Zhu, Qian, et al. EHR based genetic testing knowledge Base (iGTKB) Development. *BMC Medical Informatics and decision making* **15** (2015), 1.
- [10] Bodenreider, Olivier. The unified medical language system (UMLS): Integrating biomedical terminology. *Nucleic acids research* **32**, suppl 1 (2004): D267-D270.
- [11] Aronson, Alan R. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. *Proceedings of the AMIA Symposium*. American Medical Informatics Association, 2001.
- [12] Chen, Chih-Chiang, and Chang-Lin Chen. Clinical and histopathologic findings of superficial basal cell carcinoma: a comparison with other basal cell carcinoma subtypes. *Journal of the Chinese Medical Association* **69** (2006), 364-371.
- [13] Linos, Eleni, et al. Increasing burden of melanoma in the United States. *Journal of Investigative Dermatology* **129** (2009), 1666-1674.
- [14] Frank, Robert N., et al. Race, iris color, and age-related macular degeneration. *Transactions of the American Ophthalmological Society* **98** (2000), 109.
- [15] Whiting, David A., and Elise A. Olsen. Central centrifugal cicatricial alopecia. *Dermatologic Therapy* **21** (2008), 268-278.
- [16] Maulik, Pallab K., and Catherine K. Harbour. Epidemiology of intellectual disability. *International Encyclopedia of Rehabilitation*. 2010
- [17] Available at: <http://www.healthline.com/health/wilsons-disease#Overview1>
- [18] Schaefer, Mark, et al. Increased prevalence of subcutaneous lipomas in patients with Wilson disease. *Journal of clinical gastroenterology* **49** (2015), e61-e63.
- [19] Zhvania, M., et al. Wilson disease with thrombocytopenia (case report). *Georgian Medical News* **237** (2014), 61-64.
- [20] Cote, Roger A., et al. Progress in medical information management: Systematized nomenclature of medicine (SNOMED). *JAMA* **243** (1980), 756-762.

Address for correspondence

Hongfang Liu, Ph.D.
Mayo Clinic
Harwick Building, 205 3rd Ave SW, Rochester, MN 55905
Liu.Hongfang@mayo.edu

The Study of Pulse Condition of Traditional Chinese Medicine in Information Age

Cheng Shi^a, Yu Zhang^b, Qingyue Gong^a, Jinhai Zhou^a, Lisheng Xu^{b,c}

^a College of Information Technology, Nanjing University of TCM Nanjing, China

^b Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang, China

^c Key Laboratory of Medical Image Computing (Northeastern University), Ministry of Education, China

Abstract

Using the study of pulse condition in Traditional Chinese Medicine (TCM) as a case example, this paper discusses the characteristics of pulse information obtained by Chinese medicine practitioners, how to correctly understand the relationship between the human body and Chinese medicine in the information age, how to deal with pulse data, and how to study TCM pulse condition. Furthermore, we point out that the application of modern big data processing technology to the pulse of Chinese medicine offers new opportunities.

Keywords:

Traditional Chinese Medicine; Delivery of Health Care; Pulse condition

Introduction

In the information age, with the development of science and technology, our understanding of the human body is being gradually deepened. The observation of the human body has extended from human vision to the microscopic world of cells, proteins and genes. Medical equipment is being constantly developed, from the optical microscope to the computerized electron microscope, from the stethoscope to X-ray, CT, MRI and beyond. Surgical treatment has been expanded from operations by surgeons to remotely operated robots. Similarly, with the advent of the information age, Chinese medicine treatment and diagnostic technology is also being constantly developed, evolving from the traditional acupuncture using silver needle to electro-acupuncture, intermediate frequency treatment and laser acupuncture; from the observation of the tongue by naked eye to the acquisition and quantitative analysis on tongue images; from the fuzzy sense of pulse-taking by fingers to pulse condition acquisition and quantitative analysis; from handwritten medical records to electronic medical records.

Currently, some of the studies of traditional Chinese pulse diagnosis return to the anatomy of the cardiovascular system, such as using Wrist-type self-help sphygmomanometer device to evaluate the function of cardiovascular system. We cannot say that these results are not related to the clinical pulse image of Chinese medicine, but the relevance is not believed to be of a great degree. Pulse images within traditional Chinese medicine (TCM) should be associated with the 28 clinical pulse patterns of TCM. The fingertip's feeling during clinic pulse-taking should be analyzed and the pulse condition information should be acquired and quantified using some kinds of sensors. Following this acquisition and quantification, the fuzzy boundary between these objective data and clinical pulse measurements can be determined by big data analysis. In addition to the dynamic feeling of pulse-taking, many properties can constitute a feeling sensation such as thickness, float, length, frequency, rhythm, hard and soft, etc. In clinic, different pulse patterns can appear in one pulse image. Even for the TCM doctor with many years of experience, it is

difficult to accurately differentiate pulse patterns. Therefore, for the analysis of pulse condition by using big data, ambiguity is a big challenge.

In a word, TCM diagnosis and analysis technology is developing rapidly in the information age. However, compared to other diagnosis and treatment technologies, the development of TCM diagnosis and treatment technology is much slower. New diagnostic techniques based on TCM are not easily used in clinical practice, for example the electropulsograph and tongue image. The reason is related to the characteristics of clinical information obtained by Chinese medicine practitioners, their methods of clinical information processing and their means of treating diseases. Furthermore, one can argue that this generalization is related to the information characteristics of Chinese culture and the theory of TCM. The relationship between the human body and Chinese medicine needs to be correctly understood in the information age. Studying the pulse image as an example, this paper discusses the characteristics of pulse information obtained by Chinese medicine practitioners, how to deal with pulse data, and how to study pulse diagnosis in TCM [1].

The human body needs to be correctly recognized in information age

As we know, the human body is extremely complex. The system can use sensitive sensory organs at any time and any place to receive all kinds of information from the outside world such as visual observation, smelling, listening, and palpation. In order to maintain equilibrium, the human body ingests a variety of substances and energy through the digestive and respiratory systems, and eliminates waste by exhaling carbon dioxide, feces, urine, and sweat. In order to adapt to the changing natural and social environment, it needs to store energy, accumulate experience, learn knowledge, and update information constantly in the struggle for survival.

As can be seen from modern science, the internal environment of the human body is extremely complex. The body system is composed of a number of interrelated and mutually cooperative subsystems such as the digestive system, respiratory system, cardiovascular circulatory system, and nervous system. These systems are composed of a variety of cells and intercellular substances, which are invisible to the naked eye. From in-depth study, the human body is a huge, open, adaptive, self-organizing, complex information system. It becomes more and more complex when one examines how micro-systems communicate, transfer information, and maintain the normal activities of the human body. As a complex system, the human body is not easy to investigate.

This system is observed to have the properties of automatic regulation and self-recovery, with a complete set of defenses, such as adaptation, immunity, compensation, self-organization. In the future, if the human body incurs disease, doctors should be able to mobilize the body's own disease prevention and treatment system to the target, and it will be a big step forward in medicine [1].

TCM needs to be correctly recognized in the information age

The human body is an extremely complex information system. For human beings, the "human system" is still a "black box." In response to such a "black box", modern medicine attempts to gradually open the "black box" through our knowledge of nerves, body fluids, organs, tissues, cells, proteins, and genes. Although TCM does not clearly understand the internal structure of the human body, the collection of data still remains the basis of four diagnostic techniques: "observe, smell, ask, pulse-feeling and palpation". The data collected by these techniques reflect the disease state of the body. Using this information, treatments such as TCM tinctures and acupuncture can be performed to mobilize the body's defenses against diseases. This applies a black box control theory to the management of diseases. This method is the most practical method to solve complex problems. For thousands of years, the Chinese people have been using this principle to analyze and treat with human disease, and have created a wealth of information and methods to treat the complex system of the human body. This method of TCM which mobilizes the immune system to treat disease, is the direction of the development of modern medicine for diseases such as cancer. Therefore, the abundant clinical history of TCM is worthy of our exploration and promotion [2].

The Significance of Pulse Information of Traditional Chinese Medicine

Pulse information is a very important component of clinical information for TCM. Traditional pulse information is collected by the fingertips touching the radial artery resting against bone. This segment of the radial artery is located above the radial bone, and the tissue covering it is not thick, so it is easy to detect. Placement of the radial head also supports easy measurement of pulse. The position of the radial artery is between the aorta and the arterioles. Cardiac output produces a reflected wave of blood due to the narrowing in front of the vascular wall. These reflected waves and forward pulse waves are superimposed on the radial artery to form a pulse wave, carrying a wealth of information on the dynamic characteristics of the cardiovascular system including heart, internal organs, the end of blood vessels, etc. The circulatory system coordinates body tissue energy metabolism, oxygen and nutrients, and helps remove metabolic waste. Generally, modeling disease information from the circulatory system is very important, as it provides a basis for life. Therefore, the pulse diagnosis in TCM is a meaningful and important means of diagnosis of disease.

Chinese medicine is shrouded in mystery. However, as early as 1979, Professor Zhaorong Liu at Fudan University used numerical simulation to uncover the mystery behind slippery pulse. They simulated the slippery pulse of pregnant women by increasing the flow from the abdominal cavity; the results of the study are consistent with the clinical practice of traditional Chinese medicine [3].

Perplexity in the study of TCM pulse condition

TCM pulse condition is a visualized expression which is related to the radial pulse and the state of the patient. It can be obtained by the clinical doctor using the sensation of touch. It is an important basis for clinical differentiation of TCM syndromes. Generally speaking, it can be classified according to the descriptive adjectives of Float, Sink, Slow, Fast, Thick, Thin, Strong, Weak, Fluent, Astringent and Rhythm.[4]

In the Jin Dynasty (266-420 BCE), Shuhe Wang subdivided the pulse into twenty-four kinds in the book "*Maijing*". In the Ming Dynasty (1368-1644 BCE), Shenzhen Li described pulse with up to twenty-seven kinds in the book "*BinHu*

sphygmology". This knowledge about the pulse is in the form of natural language to provide a vivid expression for pulse diagnosis in TCM. For example, in the description of the pulse image of TCM, slippery pulse (Hua Mai) which is fluent should refer to the smooth, such as beads moving on the disk. Floating pulse (Fu Mai) refers to that which beats easily, heavily presses down and is diminished but not empty. In the understanding of diseases, floating pulse is mainly for surface symptoms; however slippery pulse is mainly for phlegm and food stagnation. Slippery pulse is also common phenomenon when women have no disease and pregnancy.

Clinically, TCM determines the pulse properties according to the comprehensive sensations from the doctors' fingertips, and combines the pulse information with patients' overall characteristics rather than understanding the pulse according to the position of modern radial artery in the cardiovascular system. For example, Yin and Yang, surface and inside, cold and heat, the actual situation of the internal organs and so on. That is, the pulse of traditional Chinese medicine is fuzzy and subjective data, and the rules of TCM treatments are abstract. These problems have caused a lot of confusion in Chinese medicine research [5].

How to Study TCM Pulse in Information age

Reviewing the history of study of the TCM concept of pulse, this study involves a lot of cross disciplinarity. In order to explore the mechanism of pulse generation, it is not only necessary to have knowledge of cardiovascular pathophysiology, but also knowledge of biomechanics, cardiovascular dynamics and blood rheology. In order to objectively reflect the information of doctors' fingers, not only reliable and sensitive sensors are needed, but further analysis is also needed as to the mechanical characteristics of the finger and the characteristics of the sensory cells of the finger surface when the doctor of traditional Chinese medicine is taking a pulse. Pulse information refers to a feeling of the doctor's finger surface, and is a doctor's subjective qualitative data. In addition, the study of pulse condition based on the application of pulse instrument in a previous study in China, has lead to doctors merely using the obtained information to analyze the cause of the disease instead of extending such information to the dynamic characteristics of the whole human body. Clinically, traditional Chinese medicine uses the information of pulse to judge overall dynamic trends of the human body such as Yin and Yang, inside and outside, cold and heat, asthenia and sthenia and so on. In that case doctors can use the theory of Chinese medicine to treat patients with syndrome differentiation. All of the above adds great difficulty to research into TCM pulse.

In the era of big data, Google Flu Trends project has been developed to predict flu using search words from people on the Internet who were seeking to investigate their health problems. Google collects people's "cough", "fever", "pain" and other subjective data determine the direction of the spread of influenza with big data analysis. This shows that the subjective data are useful as long as the correct use of big data association rules. However, there is no intersection among the "cough", "fever", "pain" and other subjective data concepts mentioned above, which results in an independent variable for big data processing.

Chinese clinical medicine, however, is based on the fuzzy sense of doctors' fingers to obtain the pulse information, and the concept of pulse is a comprehensive qualitative conception, of which the most common is composite pulse. Composite pulse is assembled by a variety of pulse characteristics. From the same pulse condition, doctors' subjective judgments result in multiple diagnoses. There therefore are intersections between the pulse data, which result in data traps for big data processing. Therefore, in order to

avoid these data traps, the pulse data must be preprocessed. First of all, it is necessary to set up a pulse acquisition and analysis instrument that simulates the pulse diagnosis of TCM and acquires the pulse condition with floating, sinking, thick, thin, strong, weak, slippery, astringent, frequency and rhythm information [3]. At present, the three-inch-foot pulse instrument developed by Shanghai University of Traditional Chinese Medicine has been developed in line with these requirements. Then, we can combine the objective data recorded by the clinical pulse condition with the characteristics of the TCM pulse condition to establish a qualitative classification database meets the requirements for the method of big data processing. There are thousands of Chinese medicine hospitals in China, in which many hospitals have adopted the "electronic medical record system" to record clinical data. If these hospitals can be equipped with pulse acquisition and analysis devices, the amount of data could be amazing. Therefore, modern large-scale data processing technology has brought new opportunities to researching pulse condition within TCM [6].

Conclusions

In information age, with the high degree of scientific and technological development, the research on the human body has extended from the human visual sense to the microscopic range such as cell, protein, gene and etc. In clinical treatment, CT, MRI, ultrasound and other equipment have been applied to detect the internal state of the human body. Also people have used a fully automated instrument to detect blood and urine and obtain patients' information. However, traditional Chinese medicine seems to have not yet entered the information age. TCM still uses sensory organs to subjectively collect clinical data and apply experience and traditional Chinese medicine theory to deal with data.

This paper analyzes the causes of this confusion and believes that the pulse data which perceived by TCM fingers are consisted of a variety of physical quantities. The obtained information by doctors' sensory organs is subjective and the boundary is fuzzy.

The clinical TCM pulse condition is composed of largely a composite pulse and doctors have different subjective judgments. There are intersections between the classification of Chinese pulse conditions, these intersections create barriers to the use of big data methods. To solve these problems, the authors suggest collecting the objective data about floating, sinking, fast, thick, thin, strong, weak, smooth, astringent, frequency, and rhythm by a pulse analysis system used for simulating the clinical Chinese medicine pulse diagnosis. Further studies on the function of Chinese medicine pulse condition within TCM analysis of diseases should be subjected to big data analysis to advance the theory of pulse condition. There are new breakthroughs in the field of Chinese pulse condition in the near future.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 61374015 and 61202258), the Fundamental Research Funds for the Central Universities (Nos. N161904002 and N130404016).

References

- [1] C. Shi, QF. Shi, "How to Understand the Traditional Chinese Medicine in the Information age," *China News of Traditional Chinese Medicine*, 2011-01-26, Science and Education Page 3
- [2] C. Shi, "Present Situation and Characteristics of Traditional Chinese Medical Information," *The Ninth International Annual Conference on Medical Information*, Seoul Korea, 1988.08, 841-844

- [3] C. Shi, "Talking about TCM pulse research," *The Sixth Annual Conference on China Biomedical Engineering*, Wuhan, 2004.04
- [4] C. Shi, KC. Luo, "Application of Biomechanics in the Exploration of TCM Pulse," *Guizhou Science*, 1983.02, 66-71
- [5] C. Shi, "Some Characteristics of Pulse Study of Traditional Chinese Medicine," *Journal of New Medicine*, 1978.03, pp. 105-106
- [6] B.Y. Liu, "Discussion on the Traditional Chinese Medicine Records System and its Particularity," *Medical Information*, 2004.01.

Address for correspondence

Lisheng Xu, Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang City, Liaoning Province, 110819 China, xuls@bmie.neu.edu.cn

Automated Evaluation System for Human Pupillary Behavior

Cleyton Rafael Gomes Silva^a, Cristhiane Gonçalves^b, Eduardo N. R. Camilo^c, Fabio Boaretti dos Santos^c, Joyce Siqueira^a, Eduardo Simões de Albuquerque^a, Fabrizzio Alphonsus Alves de Melo Nunes Soares^a, Leandro Luís Galdino de Oliveira^a, Ronaldo Martins da Costa^a

^aInstitute of Informatics, Federal University of Goiás, Goiania, Goiás, Brazil,

^bInstitute of Physics, Federal University of Goiás, Goiania, Goiás, Brazil

^cGoiás Eye Bank Hospital, Goiania, Goiás, Brazil

Abstract

Analyzing human pupillary behavior is a non-invasive method for evaluating neurological activity. This method contributes to the medical field because changes in pupillary behavior can be correlated with several health conditions such as Parkinson, Alzheimer, autism and diabetes. Analyzing human pupillary behavior is simple and low-cost, and may be used as a complementary diagnosis. Therefore, this work aims to develop an automated system to evaluate human pupillary behavior. The solution consists of a portable recording device, a pupillometer; integrated with a recording and evaluation software based on computer vision. The system is able to stimulate, record, measure and extract relevant features of human pupillary behavior. The results show that the proposed system is fast and accurate, and can be used as an assessment tool for real and extensive clinical practice and research.

Keywords:

Pupil; Computers

Introduction

Human pupillary behavior has been an increasingly important subject in scientific research, showing promising contributions to biometrics [1] and mostly to the medical field [2,3]. Controlled by the autonomic nervous system to provide more accurate vision and ensure retinal integrity [4], this behavior is considered a non-invasive and alternative method to assess neurological activity [3]. Abnormal pupillary behavior can indicate diseases and neurological disorders such as Parkinson [5], Alzheimer [6], autism [7] and diabetes [8].

This pupillary behavior is based on two reflexes, contraction and dilation, or miosis and mydriasis; both are involuntary reactions and are triggered to manage the amount of light entering the eyes [4]. Pupil size management depends on the illuminance level to which the eyes are exposed. Pupil size is inversely proportional to illuminance level. Contraction occurs as the illuminance level increases [9] while dilation occurs whenever illuminance level decreases [10].

As an important biosignal and potential indicator of health states, pupillary behavior is often evaluated by measuring pupil diameter over time under light stimuli, a process known as Pupillometry [11]. This method usually depends on modern technology composed by devices equipped with infrared cameras, which is able to record images even in low light conditions. Such devices, combined with software solutions

based on computer vision, are responsible for image acquisition, processing and feature extraction, essential steps for research involving monitoring pupillary behavior.

In this scenario, automated pupillometry solutions can provide effective technology to extract reliable data for patient medical evaluation. Previous studies have shown advances, presenting recording devices in many formats, from robust and large equipment [12] to solutions made with regular materials, as a common eyeglass frame [13]. However, the proposals are not always viable or portable devices with fully integrated systems for real and extensive clinical and research practice.

Therefore, this work aims to develop a portable and comfortable recording device, a pupillometer; with an automated and integrated evaluation system based on computer vision to perform pupillometry procedures. These two components combined form an Automated Pupillometry System (APS), capable of stimulating, recording, measuring and extracting relevant human pupillary behavior features. Results demonstrate the system as a useful and efficient solution and tool for medical and research investigations.

Methods

The method is presented accordingly in the Automated Pupillometry System (APS) stages (Figure 1).

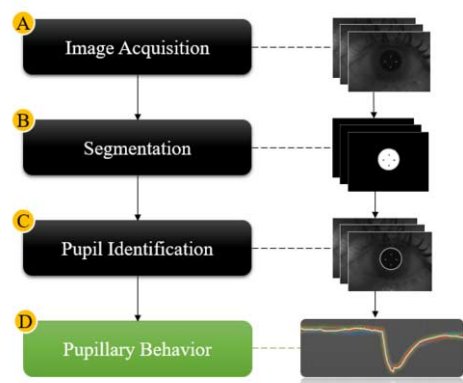


Figure 1—Stages of the Automated Pupillometry System (APS): (A) Image Acquisition; (B) Segmentation; (C) Object Identification and (D) Pupillary Behavior.

The first subsection, Image Acquisition, describes the pupilometer resources and operation, recording function of the software, participants description and the pupillometry protocol applied. The second subsection, Segmentation and Pupil Identification, presents the image processing to highlight the pupil in images and explains the measurement of pupillary diameter. The third subsection, Pupillary Behavior, shows the feature extraction.

Image Acquisition

Pupillometer

The recording device (Figure 2) was built by adapting a virtual reality glasses frame. Two infrared cameras were attached to two acrylic plates, and then on a printed circuit board with an eyeglass shape. The cameras are Point Grey's mono firefly, FMVU-03MTM-CS model, CMOS sensor (global shutter) and have a USB 2.0 interface. The circuit board was configured to control five RGB LEDs and four infrared LEDs on each side. Infrared LEDs provide illumination for a low light condition whereas RGB LEDs are used for stimuli.



Figure 2– (A) Pupillometer Off; (B) Pupillometer On.

On each side, Infrared LEDs were arranged in a rhombus format, Figure 2 (B), from five RGB LEDs; four were arranged in a square form and the last one fixed below the acrylic plate. These plates have a circle drawn and become visible by turning the lower LEDs on. Used to guide the gaze, the orientation circle's luminosity is adjusted so the patient can locate them without triggering pupillary reflexes. During the procedures, to avoid distraction, this circle is lit only in the non-stimulated side.

Recording Interface

The developed software was built in Matlab® R2016b platform, using the following toolboxes: image acquisition, image processing, parallel computing and Matlab GUI (graphical user interface). The first window is dedicated to the recording procedures (Figure 3) and has six operation modes, separated by a stimulated and a recorded side, as follows:

1. Stimulated Eye: Right; Recorded Eye: Right.
2. Stimulated Eye: Left; Recorded Eye: Right.
3. Stimulated Eye: Right; Recorded Eye: Both.
4. Stimulated Eye: Left; Recorded Eye: Left.
5. Stimulated Eye: Right; Recorded Eye: Left.
6. Stimulated Eye: Left; Recorded Eye: Both.

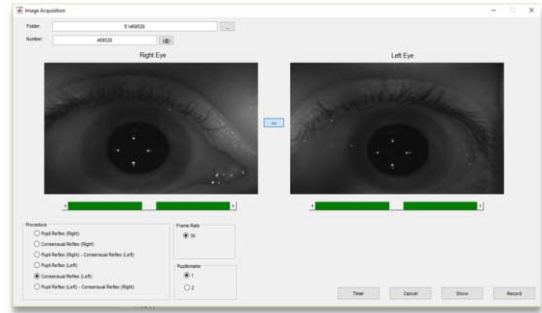


Figure 3– APS Record Window.

In addition to the direct reflection, which is the result of the behavior of the stimulated eye, options: 2, 3, 5 and 6 were created to evaluate consensual reflex, because both pupillary reflexes are expected to be symmetrical, regardless of which side is illuminated [4,14]. This particularity has a medical value as asymmetries are an indicator of optic nerve and retinal diseases [14], as well as brain trauma [4].

In this part of the system, fields and buttons allow users to: define the destination folder and the patient's ID; display the camera feeds; initiate the timer for dark adaptation; and start or abort recording procedures. The videos are recorded at 30 fps with 640x480 pixels resolution (Figure 5 (A)) and exported in an uncompressed AVI file with the name formed by the patient's ID plus the letter 'R' or 'L', depending on the recorded side.

Participants

All experimental procedures were reviewed and approved by the Research Ethics Committee in a project submitted to the Brazil platform, under CAAE number: 57179216.8.0000.0033. Forty volunteers were recruited and agreed to participate in this research by signing a free and informed consent term. The research aim and details were explained to all.

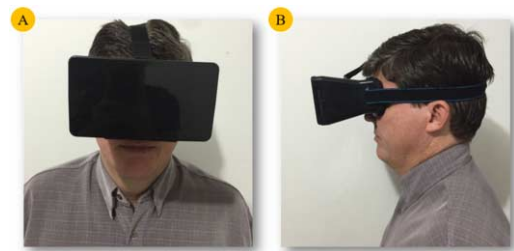


Figure 4– The pupillometer being used by a volunteer: (A) Front View; (B) Side View.

For all cases, the fifth operation mode was selected. As the pupillometer design isolates most of the external illumination, as demonstrated in Figure 4, the infrared LEDs are switched on at the beginning of the software execution (Figure 3) as well as the orientation circle. The volunteers were advised to avoid blinking and to fix the gaze guided by the orientation circle. Then the pupillometry protocol was applied, as follows.

Pupillometry Protocol

The pupillometry protocol (Figure 5) was designed to satisfy three goals: a dark adaption period of 10 minutes to reach maximum dilation [15], stimuli in colors used for medical [16,17] and biometrics [1] research, and lightless intervals between stimuli to allow the recovery of pupillary size. All these goals were combined in 14:05 minutes.

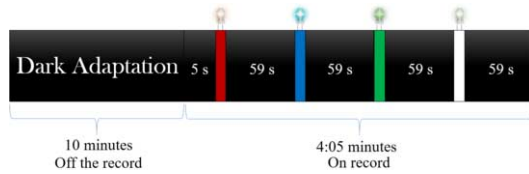


Figure 5– Pupillometry Protocol.

Procedures begin with pupillary adaptation in a room with lights off, <1 lux. After the pupillometer is put in position (Figure 4), the software starts recording the first five seconds to register dilation and continues with a cycle of four stimuli with a 1 second duration each, every 59 seconds, in red-blue-green-white order. All stimuli were set in scotopic condition (100 lux) and the intensities were checked by a TASI-8721 light meter, with $\pm 4\%$ rdg ± 10 dgts ($< 10,000$ lux) of accuracy. Infrared LEDs were adjusted considering the limit for the eyes' exposure [18] to 4:05 minutes of recording procedure.

Segmentation and Pupil Identification

Through segmentation, the pupillary region can be isolated and measured in the images. The end result is a signal of the pupillary diameter as a function of time, representing the patient pupillary behavior during the procedure. This signal gives rise to relevant features used to evaluate the pupil behavior. In this work the segmentation proposal consists of six steps (Figure 6).

Gaussian Filtering

The first step, Figure 6 (A), eliminates noise in images. Noise might cause interference in the subsequent processes. Noise is relatively common in the image acquisition process and to minimize the effects, smoothing techniques can be applied [19]. Thus, at this stage, the images are smoothed using a Gaussian filter with $\sigma = 2$ and window size = 9. These parameters, defined experimentally, allow most of the noise to not be highlighted in the binarization.

Binarization (Weighted Otsu)

Through the binarization, regions in the image can be highlighted in two different colors; black and white. For this division to be made, a threshold representing the intensity that divides such regions needs to be defined. Although the pupil represents a large and connected dark region in the image, finding the ideal threshold is not a trivial task. Position, color of adjacent areas, facial structure, color and intensity of stimuli, and proximity to the camera are examples of situations where image characteristics can change dramatically. In these cases the same threshold does not get a correct separation.

In this scenario it is important to adapt to these changes and to perform binarization using automatic thresholds. In this step, a segmentation similar to that used in [20] is applied, when using the Otsu's method. This method seeks to minimize the sum of the object and background variance. For this, the technique performs iteration for all possible thresholds until it finds one that minimizes the sum of the object and background variances.

This method applied to an image that contains the eye structure tends to separate the sclera region, white region, from the pupil and iris, darker regions. However, in order to restrict the pupillary region of the iris region, a weighting factor of 2.5 is applied. This weighting of the otsu value (WO) is efficient and adaptive, as can be seen in the values and results shown in Figure 6 (B). The factor was defined by means of empirical tests on images with different illuminations, proximities and ocular structures.

Morphological Filtering

During the recording procedure and due to the stimuli and the infrared illumination, videos have moments with different levels of illumination. The main difficulty is the presence of reflexes produced by the cornea. For this reason part of the pupillary region is not captured or is extremely compromised (Figure 6 (A-2)). In order to overcome this problem, morphological operations of opening and closing in conjunction with a flood-fill process to close holes are applied in the image resulting from the binarization (Figure 6 (C)).

Numeric Shape Descriptor

Once the objects are isolated in the previous process, there is the possibility that other objects may be highlighted, such as: remaining noises, occluded pupil, pupil with a large part covered by the eyelid and/or eyelashes, as can be seen in Figure 6 (C- 2), or another large dark region. A numeric shape descriptor is used to recognize the pupil or detect moments in which measurement is not possible, such as blinking occasions.

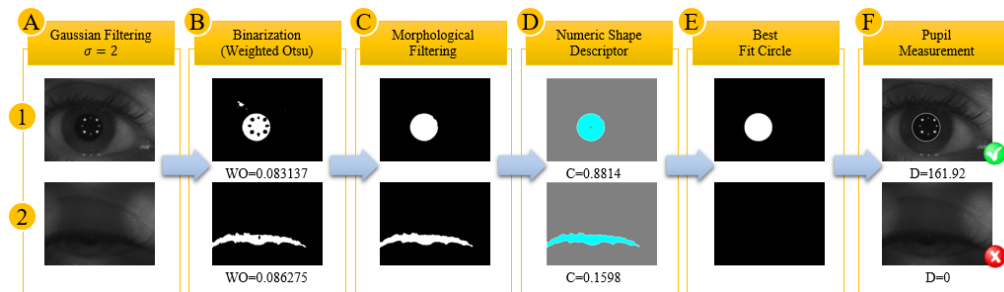


Figure 6– Segmentation: (A) Gaussian Filtering; (B) Binarization (Weighted Otsu); (C) Morphological Filtering; (D) Numeric Shape Descriptor; (E) Best Fit Circle and (F) Pupil Measurement, (1) Normal, (2) Blinking.

To recognize the pupil or detect moments that there is no possibility of measurement, such as blinking moments, a numeric shape descriptor is used.

The descriptor, equation 1, allows the circularity of the objects in the image to be calculated [19]. The value resulting from the calculation approaches 1 as the circularity increases, and the value 1 represents a perfect circle. Considering that the pupil does not represent a perfect circle and that in some moments there will be deformations resulting from reflexes, a threshold of 0.65 was experimentally defined and proved to be effective. This value allows to separate images eligible for measurement from those where there is absence of a sufficient pupillary region, as shown in Figure 6 (D-2).

$$C = \frac{4\pi * Area}{Perimeter^2} \quad (1)$$

Best Fit Circle and Measurement

Images identified with circular objects pass through a last step before they have their diameter measured. To correctly measure the diameter, even in images whose pupil is partially segmented, a circular fitting process (Figure 6 (E)) calculates the circle that best fits the rectangular limits of the object. This procedure allows for a simple and fast operation to find the approximate pupil diameter (Figure 6 (F)). The diameter extracted through this process is a good approximation of the actual diameter even in images with partial pupils. For research that aims to use the iris boundaries, where it is important to isolate the original pupil shape, the method is also efficient. In these cases the fitting step should be ignored.

Pupillary Behavior

Feature Extraction

The points of interest in the data are those closest to the stimuli. For this, the signal is decomposed by the system in four parts, one for each region influenced. Representing five seconds prior and ten subsequent to a second of stimulus, each decomposed signal represents sixteen seconds of observation. Eight features are extracted from each signal. These features are relevant in order to understand changes in human pupillary behavior [11,21]. Such features are given in Figure 7.

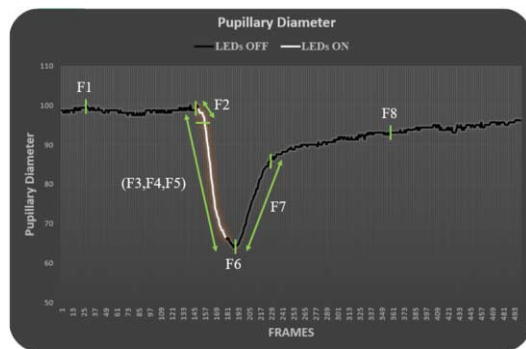


Figure 7— Features: F1- Maximum dilation; F2-Latency; F3- Time to maximum contraction; F4- Relative amplitude; F5- Absolute amplitude; F6- Maximum contraction; F7-Post-Illumination response; F8-6 sec test.

F1. Maximum dilation: maximum diameter registered in the signal, usually found in the period before the start of the stimulation.

F2. Latency: time spent in seconds between the start of the stimulus and the beginning of the process of contraction, thus considered as a 10% change in diameter.

F3. Time to maximum contraction: time spent in seconds between the start of the stimulus and the maximum contraction registered.

F4. Relative amplitude: percentage value of the difference between the largest and smallest diameter.

F5. Absolute amplitude: absolute value of the difference between the pupillary diameter before the stimulus and the value of maximum contraction.

F6. Maximum contraction: smaller diameter contraction registered, usually found during stimulation.

F7. Post-Illumination response: time in seconds the pupillary diameter takes after the stimulus to achieve 85% of the previous value.

F8. 6 sec test: diameter registered 6 seconds after the end of the stimulus.

Results and Discussion

In this research, the experiment was accomplished with 40 volunteers. In all cases both eyes were recorded with the right side receiving stimuli. As each video was recorded at 30 frames per second, they are composed of 7.350 frames each. The procedure was able to extract the pupil behavior using the diameter and calculate the features successfully in all cases. A total of 588.000 frames were processed by the system. At the end of each processing a screen displays the results, as shown in Figure 8. All data are stored automatically in an Excel file, with the diameter values in pixels and percentage scales. The file also contains a comparison between the left and right side.

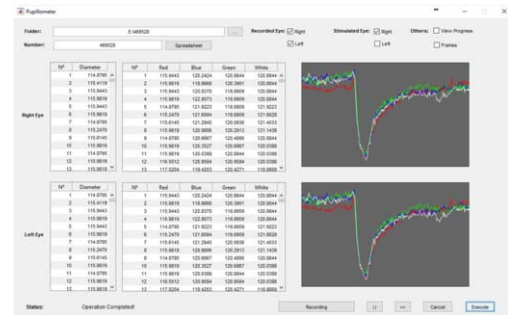


Figure 8— APS Main Window.

In order to measure the performance and effectiveness of the segmentation process, the images from each participant were separated into five categories, one for the moment without stimulation and one for each color used: red, blue, green and white. This is justified by the fact that each color of stimulus changes the features of the image differently. After this separation, two images for each period and participant were randomly selected, totaling 400 images.

Through segmentation of the selected images, the average speed processing was 347.84 milliseconds. For testing, a computer with intel core i5 processor and 4 GB of RAM was used. For the calculation of accuracy, the number of correct segmentations was divided by the total number of images. By means of this calculation, the percentage of accuracy of the approach reached the value of 97.25%.

It is worth mentioning that the images on which the method was not successful were those whose pupil was partially covered by eyelashes or eyelids. Images on which the pupil was totally or practically covered were eliminated and replaced by a new image of the same category. It is important to emphasize that the numeric shape descriptor used allowed 100% of these cases of pupillary occlusion to be automatically excluded.

Conclusion

This study proposes an Automated Pupillometry System (APS). The segmentation process is essential for this approach and is fast and effective, reaching a level of accuracy of 97.25 %. This factor allowed the system to be able to extract 8 important features of the human pupillary behavior for each type of stimulus, totaling 32 features. All the processes from the acquisition, to the generation of the data in the file were made in automated form; fulfilling the main goal of this work.

Blinking is one of the main problems in pupil segmentation and the main cause of noise in data. The usage of a shape descriptor is able to remove it in its totality. Therefore, the system represents an efficient pupillometry solution for extensive practices in medical and research investigations. In future work, the objective is to implement functions to allow the pupillometry protocol to be customized, considering the time, color and interval between stimuli. This will make procedures more flexible for even wider use.

Acknowledgements

This research is financially supported by Coordination for the Improvement of High Level Personnel - CAPES.

References

- [1] R.M. da Costa and A. Gonzaga, Dynamic Features for Iris Recognition, *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **42** (2012), 1072-1082.
- [2] K. Yamaji, Y. Hirata, and S. Usui, The pupil as a possible monitor of the autonomic nervous system, in: *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 'Magnificent Milestones and Emerging Opportunities in Medical Engineering' (Cat. No. 97CH36136)*, IEEE, 1997.
- [3] Y. Altay, M.M. Altay, G. Demirok, O. Balta, and H. Bolu, Measurements of Pupillary Diameter and Wavefront Aberrations in Pregnant Women, *Scientifica* **2016** (2016), 1-5.
- [4] D. Purves, *Neuroscience*, Sinauer Associates, Publishers, Sunderland, Mass., 2004.
- [5] C.-A. Wang, H. McInnis, D.C. Brien, G. Pari, and D.P. Munoz, Disruption of pupil size modulation correlates with voluntary motor preparation deficits in Parkinson's disease, *Neuropsychologia* **80** (2016), 176-184.
- [6] F. Fotiou, K.N. Fountoulakis, M. Tsolaki, A. Goulas, and A. Palikaras, Changes in pupil reaction to light in Alzheimer's disease patients: a preliminary report, *International Journal of Psychophysiology* **37** (2000), 111-120.
- [7] H.J. Nuske, G. Vivanti, K. Hudry, and C. Dissanayake, Pupillometry reveals reduced unconscious emotional reactivity in autism, *Biological Psychology* **101** (2014), 24-35.
- [8] K. Hayashi and H. Hayashi, Pupil size before and after phacoemulsification in nondiabetic and diabetic patients, *Journal of Cataract and Refractive Surgery* **30** (2004), 2543-2550.
- [9] T.M. Gable, A.L. Kun, B.N. Walker, and R.J. Winton, Comparing heart rate and pupil size as objective measures of workload in the driving context, in: *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, ACM Press, 2015.
- [10] C.-A. Wang and D.P. Munoz, A circuit for pupil orienting responses: implications for cognitive modulation of pupil size, *Current Opinion in Neurobiology* **33** (2015), 134-140.
- [11] H. M. Pinheiro, E. Nery, R. Camilo, M. Oftalmologista, and R. Martins Da Costa, Metodologia e Dispositivo Portátil para Avaliação do Re-

- flexo Pupilar, in: *21st Americas Conference on Information Systems, AMCIS 2015*, Association for Information Systems, Puerto Rico, 2015.
- [12] W. Nowak, A. Zarowska, E. Szul-Pietrzak, and M. Misiuk-Hojlo, System and measurement method for binocular pupillometry to study pupil size variability, *Biomed Eng Online* **13** (2014), 69.
- [13] Y.-C. Tsai, Y.-J. Yan, M.-L. Ko, T.-W. Huang, J.-C. Chiou, and M. Ou-Yang, Design of synchronizing pupillometer for observing nerve conduction by pupillary responses, in: *2016 IEEE International Instrumentation and Measurement Technology Conference Proceedings*, IEEE, 2016.
- [14] O. Bergamin, M.B. Zimmerman, and R.H. Kardon, Pupil light reflex in normal and diseased eyes, *Ophthalmology* **110** (2003), 106-114.
- [15] L.A.R. Gabriel, N.S. Peachey, and J.S. Sunness, Retinal Function Testing and Genetic Disease, in: *Genetic Diseases of the Eye*, Oxford University Press, 2012, pp. 343-355.
- [16] S. Traustason, A.E. Brondsted, B. Sander, and H. Lund-Andersen, Pupillary response to direct and consensual chromatic light stimuli, *Acta Ophthalmologica* **94** (2015), 65-69.
- [17] M. Bernabei, L. Rovati, L. Peretto, and R. Tinarelli, Measurement of the pupil responses induced by RGB flickering stimuli, in: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, IEEE, 2015.
- [18] E. Allen, Eye Safety for Proximity Sensing Using Infrared Light-Emitting Diodes Photobiological Effects of Application Note 1737 IEC-62471, *Photobiological* (2012), 1-11.
- [19] J.C. Russ, Image analysis and signal processing in electron microscopy. P. W. Hawkes, W. Owen Saxton, F. Peter Ottensmeyer and Azriel Rosenfeld (Editors). Published by Scanning Microscopy International, 1988, Chicago, IL, USA; ISSN 0892-953X, *X-Ray Spectrometry* **18** (1989), 247-247.
- [20] M. Abdullah, Fast and Accurate Pupil Isolation Based on Morphology and Active Contour, *International Journal of Information and Electronics Engineering* **4** (2014).
- [21] P. Adhikari, A.J. Zele, and B. Feigl, The Post-Illumination Pupil Response (PIPR), *Investigative Ophthalmology & Visual Science* **56** (2015), 3838.

Address for correspondence

Cleyton Rafael Gomes Silva, E-mail: cleytonrafael@inf.ufg.br.

Prescription Opioid Dependence in Western New York: Using Data Analytics to Find an Answer to the Opioid Epidemic

Shyamashree Sinha^a, Gale R Burstein^b, Kenneth E Leonard^c, Timothy F Murphy^d, Peter L Elkin^a

^a Department of Biomedical Informatics, University at Buffalo - SUNY, Buffalo, New York, USA

^b Erie County Department of Health, Buffalo, New York, USA

^c Research Institute on Addiction; Psychology; Psychiatry, University at Buffalo - SUNY, Buffalo, New York, USA

^d Clinical and Translational Research Center, University at Buffalo - SUNY, Buffalo, New York, USA

Abstract

Opioid dependence and overdose is on the rise. One indicator is the increasing trends of prescription buprenorphine use among patient on chronic pain medication. In addition to the New York State Department of Health's prescription drug monitoring programs and training programs for providers and first responders to detect and treat a narcotic overdose, further examination of the population may provide important information for multidisciplinary interventions to address this epidemic. This paper uses an observational database with a Natural Language Processing (NLP) based Not Only Structured Query Language architecture to examine Electronic Health Record (EHR) data at a regional level to study the trends of prescription opioid dependence. We aim to help prioritize interventions in vulnerable population subgroups. This study provides a report of the demographic patterns of opioid dependent patients in Western New York using High Throughput Phenotyping NLP of EHR data.

Keywords:

Substance-Related Disorders; Prescriptions; Natural Language Processing

Introduction

Dependence and abuse of prescription opioid pain medication has substantially increased over the last decade. The rise in opioid dependence contributes to the rising prescription drug overdose deaths over the last decade. In 2014, approximately 1.3 million adults aged 26 or older had a pain medication use disorder in the past year. According to the 2014 National Survey on Drug Use and Health, published by the Substance Abuse and Mental Health Service Administration (SAMSA) about 0.6 percent of the population aged 26 or older were opioid dependent [1].

New York State has seen a steady rise in opioid related deaths from 1 per 100,000 people in 2010 to 4.2 per 100,000 people in 2013 [2]. This is considerably higher than many other states in the United States. The state health department is taking steps towards controlling this epidemic by training first responders, fire fighters and local health officials to recognize and treat opioid related overdoses. The non-medical use of prescription drugs has been on the rise among persons 12 years and older. This has led to the challenge of clinicians treating patients with chronic pain and the decision to treat the pain with opioid medications.

The New York State Department of Health (NYSDOH) implemented a prescription drug monitoring program, I-

STOP, in August 2013. This gives the prescribing providers a secure access to the patients' registry of class II, III and IV controlled substance medication, which they are expected to consult before ordering any prescription pain medications. In 2016 NYSDOH implemented an amendment to the New York State Public Health Law §3331, that limits the prescription of controlled substances duration to seven days for the first prescription of narcotics for an acute pain condition [3].

There are many studies on the risk of opioid abuse and overdose among opioid dependent people [4, 5]. In March 2016, the Centers for Disease Control and Prevention (CDC) published chronic pain medications prescribing guidelines which presented an evidence based approach to patient assessment and prescription opioid monitoring tools [6].

On reviewing recent clinical evidence with an aim to propose adequate guidelines for prescription of pain medication the CDC concluded that there is insufficient evidence of effectiveness of long term opioid use for chronic pain [7]. In addition there is evidence of increased risk of overdose among patients using Methadone for long term opioid therapy or dependence [8]. Long term prescription opioid therapy for acute pain indications may result in higher incidence of opioid dependence [9, 10].

Knowledge about the epidemiology of local opioid dependent people may enable health care providers and population health workers to take proper actions. We used our web server based Natural Language processing (NLP) system to identify the cases of opioid dependence among our primary care clinic population in the Western New York (WNY) area.

The study of the distribution and determinants of opioid dependence among patients who are treated with chronic pain medications prescribed by their healthcare providers would aid in answering some key questions about potential abuse and overdose on opioids. The descriptive epidemiology of opioid dependence would help in identifying vulnerable age groups, gender, race, ethnicity, regional distribution and type of opioid pain medications, that more commonly result in dependence.

Methods

Technology

We implemented an Observational Medical Outcomes Partnership / Observational Health Data Sciences and Informatics (OMOP / OHDSI) database, to hold structured EHR data from local area primary care clinics managed by Allscripts company. We also created a high throughput

phenotyping, NLP system, which can parse 7 million clinical notes in 1.5 hours. This runs as a web service and provides a modular component based NLP system. After the full semantic parse, we match the content against any number of ontologies. For each match, we tagged it as either a positive, negative, or uncertain assertion. We then performed automated compositional expressions. We stored the codes in a Berkeley database (BDB) NOSQL database, and the compositional expressions are stored in Neo4J (a graph database) and Graph DB (a triple store). This flexibility allows rapid retrieval of complex questions in real time.

In comparing the NOSQL database's retrieval times to SQL queries with either Oracle or SQL Server, we found the NOSQL database to improve performance between 100-1000 fold. Bio surveillance of influenza from clinical encounter notes using this NLP system has been shown to be superior to the conventional tools of surveillance, by Elkin et. al in Ann. Int. Med. 2012 [11]. Evidence of effective monitoring tools for post-operative complication in patients in the VA hospital systems with this method has also been published in JAMA, 2011 [12].

Analysis

The retrospective analysis of EHR data from local clinic patients was performed using queries on the problem list, demographic data and medication list of all the patients in the data base. The OMOP/OHDSI data base was collected from Allscripts EHRs from 2010 to 2015. This common data model helps in the systematic analysis of disparate observational data bases of clinic records from the primary care and family medicine clinics in WNY region.

The database contained 212,343 patient records that were parsed and deidentified. Specific research IDs were assigned to each of the patient records and stored in a secure firewalled device for data analytics. The entire 212,343 records were queried for opioid dependence from the ICD9 and 10 diagnostic codes and SNOMED CT codes mapped to both the clinical notes and the problem list for each patient based on the mapped ICD and SNOMED CT codes. 1356 patients were identified as to having opioid dependence. Based on the age distribution (age range of 19 to 89 years) the population was divided into eight age groups (Table 1).

Table 1 - Age distribution of study population

Age Group (in years)	Opioid Dependent	Total population
19-28	279	36465
29-38	534	26313
39-48	236	25365
49-58	170	36270
59-68	106	36825
69-78	29	25313
79-89	2	25792
Total	1356	212343

High Throughput Phenotyping

The High-Throughput Phenotyping (HTP)-NLP subsystem is a software that produces, given biomedical text, semantic annotations of the text. The semantic annotations identify conceptual entities - their attributes, the relations they have with other entities and the events they participate in, as expressed in the input text. The conceptual entities, relations, attributes, and events identified are specified by various knowledge representations as documented in Coding Sources. Examples of coding sources are medical terminologies [e.g.,

SNOMED CT, RxNorm, LOINC) and open biomedical ontologies foundry ontologies, e.g. Gene ontology, Functional Model of Anatomy]. The annotation results may be displayed or output in formats suitable for further processing. Entity identified is assigned a truth value from zero to one. Values from the text are assigned to entities from ontologies such as SNOMED CT. Where applicable data were combined into post-coordinated compositional expressions. These are fit into clinical models such as a) Course of Illness, b) Course of Treatment, and c) Course of Hospitalization. This method provides the most accurate NLP solution for health available based on published numbers. It is the only system that handles uncertainty and handles automated post-coordination which has been proven to be required for 41% of the problems with which clinicians commonly have to deal [13]. The system will be used to generate semantic triples for use in referent tracking and overlaying of clinical research models.

Results

Of the 212,343 patients in the database, 1356 patients revealed opioid dependence on the problem list, ICD9-10 codes and prescription opioid pain medication with or without Buprenorphine or Suboxone in the medication list. The prevalence of opioid dependence in the clinic population was 0.64% (95% CI:0.61%-0.67%) over a five-year period.

The highest numbers of opioid dependence were seen in the 29 to 38 years' age group. That comprised 39.38% (95% CI: 36.78% to 41.98%) of the total opioid dependent population but accounted for only 2.03% of whole clinic population in this age group (95% CI:1.86% to 2.2%) (Table 1).

The subjects were then stratified by gender, race and ethnicity. There were 1005 patients with opioid dependence, among the Non-Hispanic population (total number 108,402).

Among the white Non-Hispanic population with opioid dependence, 41.33% (95% CI: 38.27% to 44.39%) were 29-38 years old. The next common age group among the White Non-Hispanic opioid dependent subjects was 19 – 28 years, comprising of 22.48% (95% CI: 19.88% to 25.08%) of the total number of white Non-Hispanic or Latino opioid dependent population (Figure 1) (Table 2).

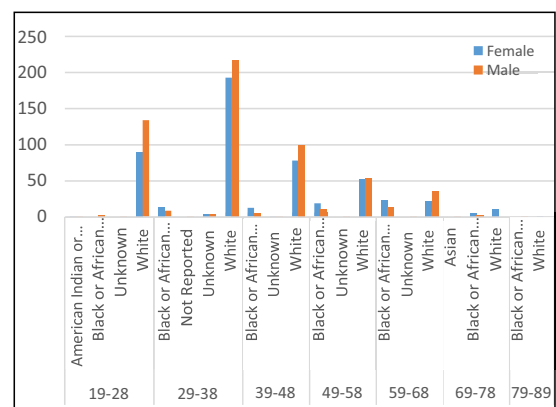


Figure 1 – Distribution of Opioid Dependence among the Non-Hispanic community in the clinic population of Western New York (x axis- age and ethnicity, y axis- number of people)

Table 2 – Distribution of opioid dependence among Non Hispanic dependent population.

	Female	Male	Grand Total
19-28	90	139	229
American Indian/ Alaska Native	1	1	2
African American		3	3
Unknown		1	1
White	89	134	223
29-38	210	231	441
African American	13	9	22
Not Reported		1	1
Unknown	4	4	8
White	193	217	410
39-48	92	106	198
African American	13	5	18
Unknown	1	1	2
White	78	100	178
49-58	71	66	137
African American	19	11	30
Unknown		1	1
White	52	54	106
59-68	46	51	97
African American	24	14	38
Unknown		1	1
White	22	36	58
69-78	16	9	25
Asian		1	1
African American	5	3	8
White	11	5	16
79-89	1	1	2
African American		1	1
White	1		1
Grand Total	526	603	1129

A total of 35/1356 patients with opioid dependence were from the Hispanic community. The rest of the 311 opioid dependent patients in the clinic did not have any race ethnicity records. The distribution of opioid dependence among the Hispanic population is shown in Figure 2.

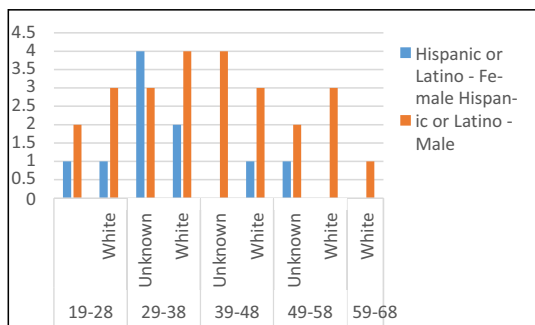


Figure 2 – Distribution of Opioid Dependence among Hispanic population (x axis- age and ethnicity, y axis- number)

We also queried the database to identify the names and quantity of opioid medication(s) used and the provider identification numbers associated with that prescriptions. The most commonly written prescription pain medication was of Hydrocodone Acetaminophen 7.5-500 mg tablets. The

providers writing the most prescriptions could also be identified based on the provider identification numbers (Table 3). This information can be useful in following providers with a higher rate of prescription for opioid pain medications. The first three numbers of the zip codes were obtained for these patients and one particular area was found to be more affected than others (Figure 3).

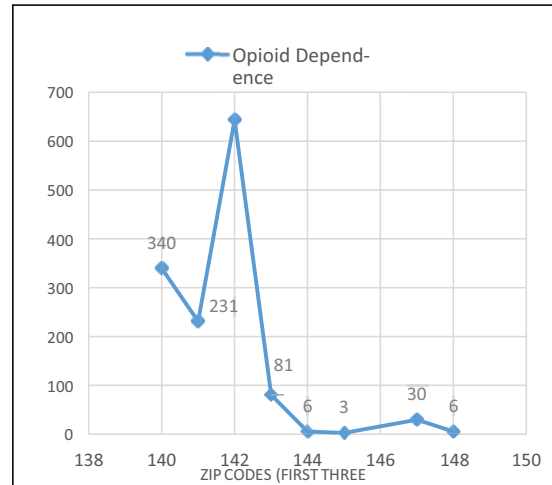


Figure 3 - Distribution of opioid dependent patients (y axis) according to first three numbers in zip codes (x axis)

Discussion

The trends of opioid dependence among the WNY clinic population in our study indicate that the prevalence is more in a certain section of the population. The predominance is among the Non-Hispanic, white population in the 19 to 38 years of age group. The prevalence in younger age implies that the complications related to opioid dependence would become a costly burden of diseases for a longer duration of time. The prevalence of dependence in this clinic population will rise if this trend continues. Interventions at curbing prescription opioid dependence is necessary in this area.

The recent CDC [7] and SAMSHA [1] reports on drug overdose related deaths all indicate the urgent need to have a multidisciplinary approach for guidelines implementation, access to medication assisted treatment programs and naloxone training and distribution to first responders. In order to prevent overdose related deaths our HTP-NLP method of knowledge extraction from EHR is useful in cost effective and timely implementation of intervention.

There is a significant body of evidence indicating that short term acute pain medication prescriptions can lead to opioid dependence [9,10]. In addition, more people die of overdoses from prescription medication than from heroin, which is a rapidly growing problem [4, 6, 8].

Observational databases with codified data from the clinical notes using standardized ontologies such as SNOMED CT can be very useful for identifying problems of high socioeconomic impact, like prescription opioid addiction. NOSQL databases can make accessing large databases with billions of values practical for real time retrieval. We have knowledge from prior studies that robust ontology based SNOMED CT has

good coverage of clinical thoughts and concepts in terms of surveillance [11, 12]. This study paves the way towards further development of real time bio surveillance with the tools of informatics.

Table 3 – Commonly written opioids

Medication	Quantity To Dispense	Activity Type Code	Medication Status	Provider number
Hydrocodone-Acetaminophen 7.5-500 MG/15ML SOLN	2250	Order	Active	168
Hydrocodone-Acetaminophen 7.5-325MG/15ML Oral Solution	1900	Order	Active	428728
Hydrocodone-Acetaminophen 7.5-325MG/15ML Oral Solution	1350	Order	Active	405425
Hydrocodone-Acetaminophen 7.5-500 MGTABS	540	Order	Active	387515
OxyCODONE HCl - 5 MG Oral Tablet	540	Order	Active	406621
Hydrocodone-Acetaminophen 7.5-500 MG/15ML SOLN	480	Order	Active	407834
Hydrocodone-Acetaminophen 7.5-325MG/15ML Oral Solution	450	Order	Active	405419
Tylenol with Codeine #3 TABS	400	Order	Complete	177
Hydrocodone-Acetaminophen 10-325MG Oral Tablet	360	Order	Complete	406623
OxyCODONE HCl - 5 MG Oral Tablet	360	Order	Active	174

There were a few limitations to this study. As it was a retrospective study on an existing database, the time to create an expert system and testing it out was not factored in. The dates and exact duration of the prescribed medications were not available. The indication for prescription and duration of symptoms were not clear. The question of potential illicit drug abuse or access and availability of methadone or buprenorphine for medication assisted treatment is not known. The potential confounders of alcoholism, smoking and or history of major depressive episodes were not taken into consideration. We assumed these are distributed uniformly over age groups, ethnicity, race and geographic distribution.

Conclusions

This was a retrospective analysis of provider EHR data from local WNY area primary care clinics looking at the epidemiology of the opioid dependent population, the rate of opioid prescriptions that may be monitored through the EHRs and the geographic distribution of the population in question.

This could provide insights into patient care patterns on a real time basis. The distribution of the opioid dependent population identified in this study indicate a disparity by age, ethnicity and race. The geographical distribution of the dependent population shows a peak in a particular region. These informations could be used to allocate resources for special preventive programs in these areas.

The young age of many of the addicted patients gives rise to the question whether legitimate opioid prescriptions are leading these age groups towards addiction. This emphasises the need for routine screening for substance abuse in patients who are on opioid pain medications. The evidence of overdose and risk of abuse with other illicit agents in this population indicate the needs for realtime intervention among opioid dependent population. Future health informatics based research should be directed towards the surveillance and intervention of patients on long term opioid pain medication.

Observational databases linked to codified NOSQL datastores powered by high throughput phenotyping approaches are a useful mechanism for warehousing translational datasets such as EHR data including codified clinical notes and reports. This mechanism can be used to implement clinical decision support tools that would provide information tailored to specific patient needs. Prior studies in this area on monitoring post operative complications within electronic medical records and biosurveillance of influenza have established evidence for this method [11,12].

Further translational studies and implementation efforts are needed in the area of opioid prescribing and addiction in order to find a cost effective and quality intensive strategy for the nationwide opioid epidemic.

Acknowledgements

Funding for this project was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001412. This study was also supported in part by an T32 grant (T32 GM099607) and an NIH NCI / VA BD-STEP Fellowship in Big Data Science. Mr. Sylvester Sakilay, a senior data analyst in the Department of Biomedical Informatics at Jacobs School of Medicine and Biomedical Informatics at University at Buffalo, State University of New York, with accessing the OMOP and high throughput phenotyping database.

References

- [1] Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. (2015). Behavioral health trends in the United States: Results from the 2014 National Survey on Drug Use and Health. Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at <http://www.samhsa.gov/data/sites/default/files/NSDUH-FRR1-2014/NSDUH-FRR1-2014.pdf> Accessed on 04/11/17
- [2] CDC, NCHS, Multiple Cause of Death on CDC WONDER Online Database, released 2015. <https://wonder.cdc.gov/mcd-icd10.html> Accessed on 04/11/17
- [3] NYS Bureau of Narcotic Enforcement, Laws and Regulations Accessed at: https://www.health.ny.gov/professionals/narcotic/laws_and_regulations/ Accessed On 04/11/17
- [4] Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose. *Ann Intern Med* 2010; 152:85–92.
- [5] Bohnert AS, Valenstein M, Bair MJ, et al. Association between opioid prescribing patterns and opioid overdose-related deaths. *JAMA* 2011; 305:1315–21.
- [6] Rudd RA, Seth P, David F, Scholl L. Increase in Drug-Involved Overdose Deaths United States, 2010-2015. *MMWR Morb Mortal Wkly Rep.* ePub; 16 December 2016. <https://www.cdc.gov/mmwr/volumes/65/wr/mm6505051e1.htm>

- [7] Dowell D, Haegerich TM, Chou R. CDC Guidelines for Prescribing Opioids for Chronic Pain in United States, 2016. *MMWR Recomm Rep* 2016;65(No. RR-1):1-49
<http://jamanetwork.com/journals/jama/fullarticle/2503508> Accessed on 04/11/17
- [8] Miller M, Barber CW, Leatherman S, et al. Prescription opioid duration of action and the risk of unintentional overdose among patients receiving opioid therapy. *JAMA Intern Med.* 2015; 175:608-15
- [9] Alam A, Gomes T, Zheng H, Mamdani MM, Jurlink DN, Bell CM. Long term analgesic use after low risk surgery. *Arch Intern Med.* 2012; 172:425-30.
- [10] Webster BS, Verma SK, Gatchel RJ. Relationship between early opioid prescribing for acute occupational low back pain and disability duration, medical costs, subsequent surgery and late opioid use. *Spine.* 2007; 32:2127-32
- [11] Elkin PL, Froehling DA, Wahner-Roedler DL, Brown SH, Bailey KR. Comparison of Natural Language Processing Biosurveillance Methods for Identifying Influenza from Encounter Notes. *Ann Intern Med.* 2012; 156:11-18. doi: 10.7326/0003-4819-156-1-201201030-00003
- [12] Murff HJ, Fitzhenry F, Matheny ME, Gentry N, Kottler KL, Crimin K, Dittus RS, Rosen AK, Elkin PL, Brown SH, Speroff T. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA,* 2011, 306(8):848-55. PMID: 21862746
- [13] Elkin PL, Brown SH, Husser C, Bauer BA, Wahner- Roedler D, Rosenbloom ST. An evaluation of the content coverage of SNOMED-CT for clinical problem lists. *Mayo Clin Proc.* 2006 Jun 81(6):741-8. 2006 Jun

Address for correspondence

Shyamashree Sinha, MD, MPH, MBA
 Department of Biomedical Informatics
 Jacobs School of Medicine and Biomedical Sciences University at
 Buffalo SUNY
 77 Goodell Street, Suite 540
 Buffalo, NY, USA, 14203 7168884858
shyamash@buffalo.edu

Phenotyping and Visualizing Infusion-Related Reactions for Breast Cancer Patients

Deyu Sun^a, Gopal Sarda^b, Steven J. Skube^c, Anne H. Blaes^d,
Saif Khairat^f, Genevieve B. Melton^{c,e}, Rui Zhang^{c,e}

^aSchool of Statistics, ^bDepartment of Computer Science and Engineering, ^cDepartment of Surgery, ^dDepartment of Medicine,

^eInstitute for Health Informatics, University of Minnesota, MN, USA,

^fCarolina Health Informatics Program, University of North Carolina, NC, USA

Abstract

Infusion-related reactions (IRRs) are typical adverse events for breast cancer patients. Detecting IRRs and visualizing their occurrence associated with the drug treatment would potentially assist clinicians to improve patient safety and help researchers model IRRs and analyze their risk factors. We developed and evaluated a phenotyping algorithm to detect IRRs for breast cancer patients. We also designed a visualization prototype to render IRR patients' medications, lab tests and vital signs over time. By comparing with the 42 randomly selected doses that are manually labeled by a domain expert, the sensitivity, positive predictive value, specificity, and negative predictive value of the algorithms are 69%, 60%, 79%, and 85%, respectively. Using the algorithm, an incidence of 6.4% of patients and 1.8% of doses for docetaxel, 8.7% and 3.2% for doxorubicin, 10.4% and 1.2% for paclitaxel, 16.1% and 1.1% for trastuzumab were identified retrospectively. The incidences estimated are consistent with related studies.

Keywords:

Phenotype; Patient Safety; Algorithms

Introduction

Infusion-related reactions (IRRs) and hypersensitivity are adverse drug events (ADRs) that can occur during chemotherapy administration. Nearly all chemotherapeutic agents have the potential to induce an IRRs with various severity [1]; more frequent and severe IRRs are labeled in black box warnings of the package inserts. Hypersensitivity is defined as an injurious immune reaction to an antigen; in this case, a component of a chemotherapy drug. An IRR is an unexpected reaction to a drug [1]. These terms are often used interchangeably in reference to adverse events, although the definition of hypersensitivity refers to an allergic reaction to a previous exposure. Here we include infusion-caused hypersensitivities as a type of IRRs. The incidence of these reactions varies widely among reports, for example, an estimate of IRRs incidence of general medications is 2-15% [2] and an estimate of IRRs incidence of trastuzumab is 0.7-40% [3]. There is lack of estimates of IRRs incidences of other medications against breast cancer. The signs and symptoms of these adverse events are broad, often difficult to define, and include all organ systems. Examples include anaphylaxis, rash, respiratory distress, hypotension, nausea, flushing, and fatigue. The severity of these reactions is mild to severe, in rare instances leading to death [1].

Currently, the detection of IRRs is based on manual chart review [3-5]. An appropriate phenotyping algorithm not only would assist clinicians increase performance to detect IRRs, avoid information overload, and ultimately improve patient safety, but also benefit researchers in modeling and analyzing risk factors of IRRs.

Moreover, in the highly time constrained clinical care setting, visualizing patients' active medication status and associated IRRs may assist clinicians in avoiding medication errors and would also assist researchers in understanding clinical correlations. Most current visualization systems focus on showing structured data alone, without integrating additional knowledge and phenotyping algorithms, such as serious ADRs associated with medication usage. In this pilot study, we developed and evaluated a phenotyping algorithm for identifying IRRs with four first-line breast cancer drugs, namely, docetaxel, doxorubicin, paclitaxel, and trastuzumab. In addition, we further developed a visualization prototype for correlating associated IRRs with the drugs for individual patients.

Methods

As illustrated in Figure 1, the methods consist of four parts: (1) extracting IRRs from structured drug product labels; (2) collecting relevant EHR data; (3) detecting IRRs by the designed phenotyping algorithm; and (4) rendering IRRs detected by visualization prototype and evaluation.

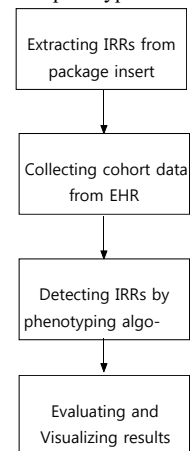


Figure 1—Flowchart of the method

In this study, we focus on the 63 brands of medications with 28 different chemical makeups approved by Food and Drug Administration (FDA) to treat breast cancer [6]. Using a previously described approach, we first extracted all IRRs (including infusion-caused hypersensitivities) from the current package inserts, i.e., FDA Structured Product Labels (FDA SPL) annotated by natural language processing (NLP) algorithms [12]. Five medications (i.e., docetaxel, doxorubicin, epirubicin, paclitaxel, and trastuzumab) among these have black box warnings of IRRs (including infusion-caused hypersensitivities) [7-11]. Since epirubicin is not found in our patient cohort, we only focus on other four medications in this study.

Clinical data for breast cancer patient cohort stored in the Fairview Epic EHR were extracted from the University of Minnesota clinical data repository. Using the domain knowledge, we categorized the IRRs into 17 symptoms according to their similarity in medicine (Table 1) and listed all the related ICD-9-CM and ICD-10-CM codes, which would be used in our phenotyping algorithm (Figure 2).

Table 1—Categories of IRRs Associated with docetaxel, doxorubicin, paclitaxel, and trastuzumab

Group Names	ICD9CM	ICM10CM
Angioedema	277.6	D84.1
Headache	339.3, 784.0	G44.41, G44.1, R51
Hypertension	405	I15.0, I15.8
Tachycardia	427.0, 427.1, 427.2, 785.0	I47.1, I47.2, I47.9, R00.0
Hypotension	458	I95.1, I95.89, I95.3, I95.2, I95.81, I95.9
Cough	519.11, 786.0, 786.2, 786.5, 786.9	J98.01, R06.9, R06.4, R06.01, R06.81, R06.3, R06.02, R06.82, R06.2, R06.00, R06.09, R06.83, R06.89, R07.9, R07.2, R07.1, R07.81, R07.82, R07.89, R05
Erythema	695	L53.0, L53.1, L53.2, L51, L52, L71.0, L71.1, L71.8, L93.0, L93.2, L49.0, L49.1, L49.2, L49.3, L49.4, L49.5, L49.6, L49.7, L49.8, L49.9, L00, L26, L30.4, L53.8, L92.0, L95.1, L98.2, L53.9
Pruritus	698.8, 698.9, 708	L29.8, L29.9, L50.0, L50.1, L50.2, L50.3, L50.4, L50.5, L50.6, L50.8, L50.9
Pain	719.4, 724.5, 724.8, 724.9	M25.50, M25.519, M25.529, M25.539, M79.643, M79.646, M25.559, M25.569, M25.579, M54.89, M54.9, M54.08, M43.8X9, M53.9
Fibromyositis	729.1	M60.9, M79.1, M79.7
Chills	780.64	R68.83
Fatigue	780.79	G93.3, R53.1, R53.81, R53.83
Dizziness	780.4	R42
Hyperhidrosis	780.8	R61
Flushing	782.62, 782.1	R23.2, R21
Nausea	787.02, 787.01, 787.03	R11.0, R11.2, R11.10, R11.11, R11.12
Miscellaneous	780.99, 995.0, 995.1, 995.2, 995.3, 995.9	A41.9, R45.84, R65.10, R65.11, R65.20, R68.89, T50.905A, T36.*X5A, T37.*X5A, T38.*X5A, T38.8*X5A, T38.9*X5A, T39.0*X5A, T39.*X5A, T39.3*X5A,

T39.95XA, T40.*X5A, T40.*05A, T40.995A, T41, T42.*X5A, T42.75XA, T43.0*X5A, T43.*X5A, T43.2*X5A, T43.5*X5A, T43.6*X5A, T43.95XA, T44.*X5A, T44.905A, T45.*X5A, T45.5*X5A, T45.6*X5A, T45.95XA, T46.*X5A, T46.9*X5A, T47.*X5A, T47.95XA, T48.*X5A, T48.2*X5A, T48.9*X5A, T49.*X5A, T50.*X5A, T50.995A, T50.A*X5A, T50.B*X5A, T50.Z*X5A, T78.3XXA, T78.2XXA, T78.4*X5A, T88.52XA, T88.59XA

As shown in Table 1, the symptoms of the IRRs are quite broad, e.g., from chills to nausea, from hypotension to fatigue. To include comprehensive information from EHR data, we used various heterogeneous data (e.g., diagnosis, lab results, and clinical notes) for IRRs detection. Collaborating with clinical experts from University of Minnesota Medical Center (UMMC), we developed a phenotyping algorithm using related concepts mentioned in the unstructured clinical notes, dynamic changes of lab tests from baseline, and related diagnosis codes (Figure 2).

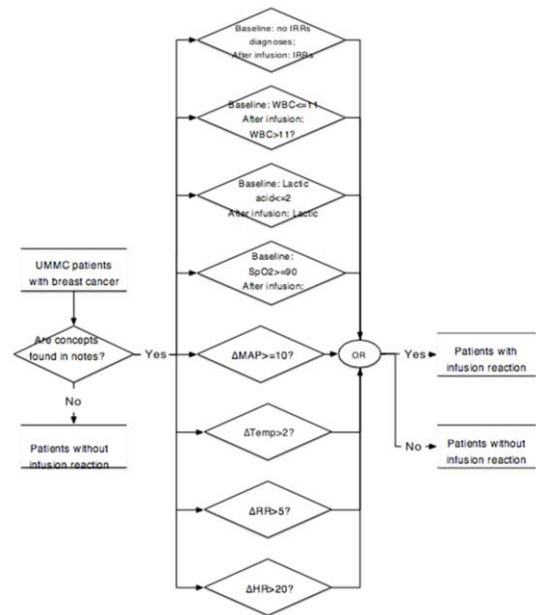


Figure 2—IRR phenotyping algorithm

Clinical notes were found containing important indicators of IRRs during the retrospective chart review by clinicians. We used open-source natural language processing (NLP) tool, BioMedICUS (BioMedical Information Collection and Understanding System) [13], to search related biomedical concepts in patients' clinical notes. BioMedICUS is an open-source NLP system based on the Unstructured Information Management Architecture – Asynchronous Scaleout (UIMA-AS) architecture specialized in NLP-related information extraction and understanding of clinical notes. BioMedICUS identifies UMLS Metathesaurus concepts (concept unique identifiers, or CUIs) from lexical variants expressed in the notes, and whether the identified concepts were used in a

negated context. As shown in Table 2, we include CUIs related to the synonyms of IRRs (Table 2) within the 24 hours of infusions of docetaxel, doxorubicin, paclitaxel, or trastuzumab from patients' clinical notes as inputs illustrated in our phenotyping algorithms (Figure 2).

Table 2– Categories of IRRs Associated with doceael, doxorubicin, paclitaxel, and trastuzumab

Group Names	UMLS CUI
Angioedema	C0002994, C0038999
Headache	C0018681
Hypertension	C1963138
Tachycardia	C0039231
Hypotension	C0020649, C3163620
Cough	C0010200, C0006266, C0008031, C0013404, C0232292, C0236071
Pruritus	C0033774
Pain	C0003862, C0004604, C1963071
Fibromyositis	C0231528
Chills	C0085593, C0026837
Fatigue	C0015672, C0023380, C0231218
Dizziness	C0012833
Hyperhidrosis	C0700590
Flushing	C0016382, C1962957, C0015230, C0221232
Nausea	C0027497, C0042963
Miscellaneous	C0020517, C0002792, C0750016, C0020538, C0042109, C0020538

In addition to clinical notes, patients with IRRs usually show abnormal lab test values, vitals, or diagnoses of abnormal syndromes. However, as shown in table 1, such syndromes are not specific to IRRs. To precisely attribute reactions to infusions, we not only apply temporal constraints (i.e. abnormal syndromes should happen within 12 hours of infusions), but also consider the change from baseline (i.e. only the syndromes that break out after infusion but not exist one month before infusion (baseline) are viewed as evidence of IRRs). The baseline of lab tests and vitals is defined as the most recent measurement within one month before infusion. The abnormal change from baseline within 12 hours can be found in two scenarios.

- First, the qualitative change, a change from a negative state at baseline to positive after infusion, e.g., a type of syndrome (diagnosed by ICD-9-CM or ICD-10- CM) shown in Table 1 that did not exist within one month before infusion but broke out right after infusion. Similar cases include WBC changes from normal ($\leq 11 \times 10^9/L$) at baseline to abnormally high after infusion ($> 11 \times 10^9/L$), lactic acid changes from normal ($\leq 2.0 \text{ mmol/L}$) to abnormally high ($> 2.0 \text{ mmol/L}$), and oxygen saturation changes from normal (≥ 90) to abnormally low (< 90).
- Secondly, quantitative change, a sufficiently large change in measurements from the baseline, e.g. the blood pressure in Mean Arterial Pressure ($\text{MAP} = [(2 \times \text{diastolic}) + \text{systolic}] / \text{divided by } 3$) with a change of no less than 10 mmHg, or a change in temperature of more than 2 Fahrenheit, or a change in respiration rate of more than 5, or a change in heart rate of more than 20 beats/min.

Only the abnormal syndromes meeting both the temporal constraints and the requirement of sufficient changes from baseline serve as evidence of IRRs. For evaluation, 42

doses of the four medications were randomly selected, the clinician then manually reviewed the charts and labeled each case as true IRR (13 cases) or false IRR (29 cases). The detection results of the phenotyping were then compared with the expert curated gold standard. The sensitivity, positive predictive value (PPV), specificity and negative predictive value (NPV) were reported. We also compared the incidence rates of IRRs on patients who used these four medications in our EHR data.

We have also created a visualization prototype to render the entire medications, vitals, labs and diagnosis data for a patient over time by using Grails framework. The interface of the visualization tool is shown and detailed in the Results section. We randomly selected two patients with true positive IRRs detected by our phenotyping algorithms and visualized their breast cancer medications and any related lab values and symptoms existed in Fairview EHR data.

Results

We applied our phenotyping algorithms into the cohort consisting of 4,084 breast cancer patients treated from the fourth quarter of 2000 to the second quarter of 2016. There are total 236,069 doses of all medications. Table 3 shows the total patients and doses of docetaxel, doxorubicin, paclitaxel, and trastuzumab. Note, epirubicin is not used in UMMC.

Table 3– Number of patients and doses on docetaxel, doxorubicin, paclitaxel, and trastuzumab in UMMC

	# of Patients	# of doses
docetaxel	188	857
doxorubicin	435	1,676
paclitaxel	576	5,918
trastuzumab	323	6,741

As an evaluation, a clinician from UMMC manually diagnosed 42 doses by chart review. Table 4 shows the detection results by our phenotyping algorithms. The sensitivity and PPV of the algorithm are 69% and 60%, respectively, while the specificity and NPV are 79% and 85%, respectively.

Table 4– Patients and doses taking docetaxel, doxorubicin, paclitaxel, and trastuzumab in UMMC

	True IRRS	False IRRS
Positive	9	6
Negative	4	23

As shown in Table 5, the prevalence of IRRs in patients who get infusions of docetaxel, doxorubicin, paclitaxel, and trastuzumab are 6.4%, 8.7%, 10.4%, and 16.1%, respectively, while the IRRs rates of all doses of docetaxel, doxorubicin, paclitaxel, and trastuzumab are 1.8%, 3.2%, 1.2, and 1.1%, respectively.

Table 5– IRRs rate among patients taking docetaxel, doxorubicin, paclitaxel, and trastuzumab in UMMC

	% of Patients	% of Doses
docetaxel	6.4	1.8

doxorubicin	8.7	3.2
paclitaxel	10.4	1.2
trastuzumab	16.1	1.1

Figure 3 renders two IRRs events. The visualization supports viewing two types of data at a time. The time axes for all the types of data are aligned. With medication information always displayed, clinicians can select one of the Vitals/Labs/Diagnosis panel as the other data source. Each medication consumption information is represented by a horizontal bar, with length of the bar reflecting the duration of consumption and height of the bar reflecting the dosage of the medicine. Since, there can be many interesting vitals/labs corresponding to a patient, the visualization provides an option to select one of the many vitals/labs available. Normal ranges for a lab/vital are also marked using the red and blue dotted lines as can be seen in Figure 3. The diagnosis data is displayed in a grid format with rows as different type of diagnosis and columns as time. A dot in a cell indicates a type of diagnosis reported on the date. The IRRs in Figure 3(a) was detected by abnormal change in blood pressure from the baseline. We can see the baseline blood pressure is 75.67mmHg measured at 9:50 am; after the patient gets infusion at 4:24 pm (labeled by the blue bar inside the block denoting a paclitaxel infusion), the blood pressure increases by 11.66mmHg, up to 87.33mmHg, which indicates an IRRs. Figure 3(b) shows another IRRs detected by the breaking-

out of a syndrome. We can see after getting infusion at 12:49 pm on 3/12/2012, the patient experienced chills and nausea on 3/13/2012; considering the clinical note of the patient also mentions the IRRs, the phenotyping algorithm classifies the reactions as IRRs.

Discussion

The incidences of the IRRs detected by our phenotyping algorithm are consistent with other researchers' studies. The incidence of four medications (docetaxel, doxorubicin, paclitaxel, and trastuzumab) varies from 6.4% to 16.1% of patients, which is consistent with the results reported (a wider estimate, 0.7%- 40% [2], and a more precise estimate 2-15% [3]). By manual chart review, L. M. Thompson et al detected 33 patients with 33 IRRs out of 197 patients with 1,788 doses of trastuzumab, resulting in an incidence of 16.2% of patients and 1.8% of doses; our method has quite close incidence to these results (16.1% of patients and 1.1% of doses).

Developing phenotyping algorithm to detect IRRs is vital for patient safety. Compared to the manual review in time-constrained clinical setting, automatic detection would greatly improve the performance of clinicians in finding

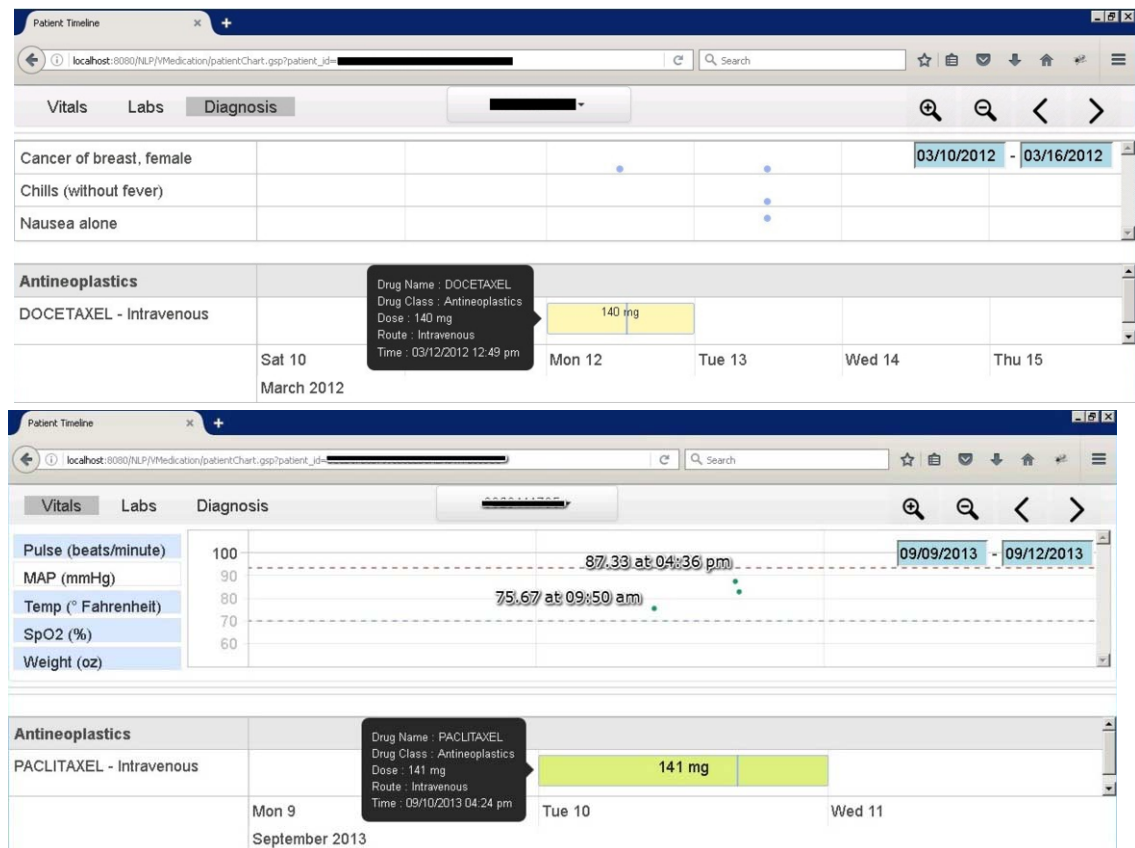


Figure 3— Visualization of two IRRs events (a) An IRR detected by greatly increased blood pressure after the infusion of 141 mg of paclitaxel. (b) An IRR of chills and nausea after the infusion of 140 mg of docetaxel.

ADRs. Accessing various types of EHR data for patients in a short time would potentially aid clinicians in making decision or changing treatment plans.

According to evaluation results, our phenotyping method has a better performance of classifying negative cases (with specificity of 79% and NPV of 85%), indicating it would be useful in reducing clinicians' workload by filtering out negative doses. On the other hand, we analyzed the relatively high false positive cases and found that most of the false positive cases are detected by changes in diagnoses (i.e. ICD codes), which is due to inaccurate time of diagnoses. The time of diagnosis recorded in the EHR database is usually one day delayed after the syndrome actually broke out (since coding takes time), as a result, some IRRs caused by previous doses were falsely identified to be caused by later doses.

Our pilot work is limited by the collection of gold standard. Since chart review is time-consuming and expensive, we only have a small size of labeled cases (42 for now) created by one physician, which limits the precision of our evaluation. Currently, the parameters of the phenotyping method are estimated based on expert's domain knowledge without referring to guidelines. In future, we are going to collaborate with more domain experts and incorporate more information, such as allergy history, and more knowledge, such as the guidelines, to automatically generate more labeled cases and improve parameter estimations by using data-driven methods.

Conclusion

In this study, we proposed a phenotyping algorithm to detect IRRs for breast cancer patients who took docetaxel, doxorubicin, paclitaxel, and trastuzumab. The phenotyping algorithm was evaluated by expert chart review. The sensitivity, PPV, specificity, and NPV are 69%, 60%, 79%, and 85%, respectively. We also developed a visualization prototype, which integrates and visualizes multiple information such as medications, vital signs, and lab test results over a specific period in one interface. The phenotyping algorithm and visualizing prototype developed in the study have potential to be adapted to detect IRRs caused by other medications. Such phenotyping algorithm and visualization method can also be applied to the clinical decision support system to improve the healthcare efficiency and patient safety.

Acknowledgements

This research was supported by the Agency for Healthcare Research & Quality grant (#1R01HS022085) (Melton), the National Center for Complementary & Integrative Health Award (R01AT009457) (Zhang), and the University of Minnesota Clinical and Translational Science Award (#8UL1TR000114) (Blazer). The authors thank Fairview Health Services for support of this research.

References

- [1] V. Kumar, A. K. Abbas, and N. Fausto, S. L. Robbins, R. S. Cotran, Robbins and Cotran pathologic basis of disease, 7th edition (2005), Philadelphia: Elsevier Saunders.
- [2] E. R. Gomes and P. Demoly, Epidemiology of hypersensitivity drug reactions, *Curr Opin Allergy Clin Immunol* 5(2005), 309-316.

- [3] L. M. Thompson, K. Eckmann, B. L. Boster, K. R. Hess, L. B. Michaud, F. J. Esteva, G. N. Hortobagyi, and C. M. Barnett, Incidence, risk factors, and management of infusion-related reactions in breast cancer patients receiving trastuzumab, *the Oncologist*, 19(2014):228-234.
- [4] W. V. Crandall and L. M. Mackner, Infusion reactions to infliximab in children and adolescents: frequency, outcome and a predictive model, *Alimentary pharmacology and therapeutics*, 17(2003): 75-84.
- [5] L. Lichtenstein, Y. Ron, S. Kivity, S. Ben-Horin, E. Israli, G. M. Fraser, I. Dotan, Y. Chowers, R. Confino-Cohen, and B. Weiss, Infliximab-related infusion reactions: systematic review, *Journal of Crohn's and Colitis*, (2015): 806-815.
- [6] Drugs approved for breast cancer. National Cancer Institute; March 2, 2016. <https://www.cancer.gov/about-cancer/treatment/drugs/breast> Accessed August 1, 2016.
- [7] Docetaxel [Package Insert]. Dr. Reddy's Laboratories Inc.; December 15, 2014. <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=45e6dce4-92e2-4ad1-bf11-bbcefb753636>. Accessed August 1, 2016.
- [8] Doxil-doxorubicin hydrochloride injection, suspension, liposomal [Package Insert]. Janssen Products, LP; March 31, 2016. <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=21d9c619-7e94-49e2-ac41-31e9ea96554a>. Accessed August 1, 2016.
- [9] Epirubicin hydrochloride-epirubicin hydrochloride injection, solution [Package Insert]. OTN Generics Inc.; June 11, 2008. <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=f7ec5ec1-fb05-4c78-b821-9aea638cf144#Prec>. Accessed August 1, 2016.
- [10] Paclitaxel-paclitaxel injection [Package Insert]. Mylan Institutional LLC; May 28, 2014. <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=9ffd3c34-537f-4f65-b00e-57c25bab3b01>. Accessed August 1, 2016.
- [11] Herceptin-trastuzumab [Package Insert]. Genentech, Inc.; March 22, 2016. <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=492dbdb2-077e-4064-bff3-372d6af0a7a2>. Accessed August 1, 2016.
- [12] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, P. Bork. A side effect resource to capture phenotypic effects of drugs, *Mol Syst Biol* 6 (2010), 1-10.
- [13] The BioMedical Information Collection and Understanding System (BioMedICUS). <http://athena.ahc.umn.edu/bio-medicus/index.jsp#/services>. Accessed August 1, 2016.

Address for correspondence

Rui Zhang, PhD, E-mail: zhan1386@umn.edu
Phone: (612)-626-8654, Mailing: MMC 912, 420 Delaware Street SE, Minneapolis, MN 55455, USA

A Deep Learning-Based Method for Similar Patient Question Retrieval in Chinese

Guo Yu Tang^a, Yuan Ni^a, Guo Tong Xie^a, Xin Li Fan^b, Yan Ling Shi^b

^a IBM Research, China, Beijing, China,

^b Pfizer, Beijing, China

Abstract

The online patient question and answering (Q&A) system, either as a website or a mobile application, attracts an increasing number of users in China. Patients will post their questions and the registered doctors then provide the corresponding answers. A large amount of questions with answers from doctors are accumulated. Instead of awaiting the response from a doctor, the newly posted question could be quickly answered by finding a semantically equivalent question from the Q&A archive. In this study, we investigated a novel deep learning based method to retrieve the similar patient question in Chinese. An unsupervised learning algorithm using deep neural network is performed on the corpus to generate the word embedding. The word embedding was then used as the input to a supervised learning algorithm using a designed deep neural network, i.e. the supervised neural attention model (SNA), to predict the similarity between two questions. The experimental results showed that our SNA method achieved $P@1=77\%$ and $P@5=84\%$, which outperformed all other compared methods.

Keywords:

Natural Language Processing; Neural Networks (Computer); Supervised Machine Learning.

Introduction

People are likely to encounter various health-related questions during their daily life. For example, patients with diabetes may want to know how to manage their diet or whether two kinds of drugs will interact. Recently, there has been a significant increase in the popularity of the online patient Q&A service worldwide. The service is provided as a website at the PC (personal computer) side or as an application on the mobile phone. In China, the most popular website of this kind is called 好大夫在线 (www.haodf.com) which has more than 130,000 registered doctors and has accumulated more than 18 million Q&A pairs; and the most popular mobile application 春雨医生 (ChunYu Doctor) has 92 million registered users and 490,000 registered doctors, and it has built up a question and answer archive of 95 million pairs. A similar situation has arisen in the U.S. as a study by Pew Internet Project's research reported: 87% of U.S. adults reported internet use, and 72% of Internet users sought health information online [1].

Given the large number of accumulated Q&A pair archive, it is likely that a newly posted question by a user would be asked by another user previously. However, the patients lack of professional knowledge or laziness for searching the archive to find a similar history question, and instead they wait for hours for the response from a doctor. To avoid the lag time and to reduce the workload on the doctors, a better method is to automatically retrieve the semantically equivalent question

from the archive. Then the answer to the equivalent question could be replied to the user immediately.

However, it is not trivial to measure the equivalence between two questions. The major challenge is the so called lexico-syntactic gap. Two questions that are semantically equivalent may be represented in very different ways. For example, both the question “我有高血脂，如果使用食疗，应该怎样配餐” (I have the hyperlipidemia, if I want to use the food therapy, what can I eat?) and the question “高血脂的病人怎样调整饮食” (How to adjust the diet for hyperlipidemia patients?) are asking the same thing, i.e. how to manage the diet for hyperlipidemia patients. However, they have few overlapping terms.

Several methods have been proposed to address the similar question retrieval problem, which could be classified into three categories. The first category of methods model this task using machine translation models [2-4]. The word-based/phrase-based translation model is trained and the similarity of two questions is measured by the probability of translating one question to another. The second category of methods use topic models [5-7]. These methods use the question corpus to train a latent topic model and the similarity of questions is measured by the similarity of mapped vectors in the latent topic space. More recently approaches explore deep neural networks. The deep learning can learn high-level feature representations by designing a deep neural network. It has achieved the state-of-art performance in many tasks. Das et al. [8] used a twin convolutional neural network with shared parameters to train the question similarity model. Lei et al. [9] designed a recurrent and convolutional model with adaptive gating to map the questions to their semantic representation. The resulting question vectors are compared with other questions via cosine similarity.

However, the above proposed approaches are applied on the general domain while there are additional challenges in handling the questions within health care domain. Firstly, the collection of training corpora is not as easy as in the general domain. For example, the word embedding used in the deep learning-based method in the general domain could use Wikipedia, while the generated word embedding is not suitable for the health care domain. Secondly, the questions in the health care domain are more complicated which involve the descriptions of diseases, signs and symptoms, etc. SimQ by Luo et al. [10] is proposed to address the retrieval of similar consumer health questions. Their method makes use of the UMLS (Unified Medical Language System) to annotate the semantic types and medical concepts from the questions. Then the similarity of questions is calculated using both semantic and syntactic features. Wongchaisuwat et al. [11] have proposed a semi-supervised method to improve the similar question retrieval and apply it in alcoholism related questions.

In this paper, we proposed to use a deep neural network model with a supervised learning to determine the similarity between patient questions. We focus on the patient questions in Chinese, which brought additional challenges. For example Chinese does not have a space to separate the word, then we need an accurate word segmentation to guarantee the final performance. We have crawled Q&A pairs from the top four online service websites. This corpus is used to generate the word embedding using via unsupervised learning. Then the word embedding is used as the input to another deep neural network leveraging supervised learning to determine the similarity between questions. The second neural network makes use of the recurrent neural network (RNN) encoder to encode the input questions and then on top of these encoders, word-by-word attention is employed to get word alignment and compare measurement for words from questions. Then an output long short-term memory (LSTM) layer is used to summarize the compare results and output the labels. A labeled dataset from human experts is used to generate the training and testing dataset for question similarity judgement.

Methods

In this section, we introduce the preparation of the corpus and how this corpus is used to generate a word embedding for patient question data. Next, we describe how we generate the training and testing dataset for the supervised learning. Finally, we illustrate the approaches we have compared and designed for similar patient question retrieval.

Word Embedding Generation

The first step of our method is to generate the word embedding for the terms in the health care domain. This involves a mathematical embedding from a space with one dimension per word to a continuous vector space with much lower dimensionality. The semantically related terms will be close to each other in the mapped space. The word embedding has been shown to be useful in various natural language processing tasks.

A large corpus that covers most of the terms in the scenario is required. In our work, we have crawled the top 4 online Q&A service websites. The obtained number of Q&A pairs (62.9 Megabytes in total) is listed in Table 1. In this work, we focus on the cardiovascular medicine related questions, thus we only crawled Q&A pairs in this area. Please note that the number of Q&A pairs on these websites is larger than what we have crawled. As these websites only expose 200-1000 pages of the most recent questions, we are not able to crawl all historical questions. But our crawler continues to crawl periodically, we will accumulate more Q&A pairs.

Table 1 – Number of Q&A Pairs

Websites	Number of Q&A Pairs
www.haodf.com	81585
www.39health.com	31173
www.120ask.com	3949
www.xywy.com	3982

After the Q&A pairs corpus is obtained, we need to pre-process the data before feeding the data to the neural network. The pre-processing includes punctuation removal and word segmentation. We use regular expressions to remove the punctuation. The word segmentation required is different with English there is no space between words in Chinese. We have tried the available word segmentation tool in Chinese.

However, the results are not satisfying as the dictionary used in the word segmentation tool does not cover enough medical terminologies. Zhang et al. [12] also proved the necessity of domain dependent word segmentors and the effectiveness of word embedding in dealing with clinical notes. Thus, we need to prepare a medical terminology to extend the dictionary. There are two challenges here; firstly, most well-defined terminologies in English do not have a Chinese version, e.g. SNOMED CT. Secondly, as the questions are from patients, the professional terminology from domain experts may not cover most of the terms from patients. In our work, we have collected the medical terms from three sources: (1) the Chinese version of the well-defined terminologies such as ICD-10 and LOINC; (2) the drugs that are crawled from the China Food and Drug Administration; (3) the drug names and the diagnoses names in the health information system. Our final dictionary contains 27322 disease names, 6021 symptom names and 5821 drug names. In addition, we have identified 33745 synonyms for these terms. We use this dictionary to extend the existing word segmentation tool and use it to do the word segmentation on the corpus.

Finally we use the Skip-Gram model in [13] to generate the word embedding. As based on Mikolov [13], the Skip-Gram is better for infrequent term representations. In our scenario, drugs, symptoms, and diseases are important terms while they may occur infrequently in the corpus. The content window size is set to 9 and the dimension of word vectors is set to 200. Table 2 shows some examples of semantically related terms captured by our word embedding.

Table 2 – Related Terms in Word Embedding

高血压 (hypertension)	冠心病 (coronary atherosclerotic heart disease)
高血压病 (hypertensive disease)	心肌梗塞 (myocardial infarction)
血压高 (high blood pressure)	缺血性心脏病 (ischemic heart disease)
高血脂 (hyperlipidemia)	心肌梗死 (myocardial infarction)
遗传性 (hereditary)	心绞痛 (angina pectoris)
高血糖 (hyperglycemia)	高脂血症 (hyperlipidaemia)
低血压 (low blood pressure)	冠状动脉硬化性心脏病 (coronary atherosclerotic heart disease)
脑血栓 (cerebral thrombosis)	心梗 (myocardial infarction)
颈椎病 (cervical spondylopathy)	心血管疾病 (cardiovascular disease)

Dataset

We invited domain experts to annotate the semantically equivalent questions. We selected 1688 questions in the hyperlipidemia area and asked the domain experts to group the semantically equivalent questions. Finally, the domain experts divided these questions into 241 groups. We randomly selected 1488 questions for training 100 questions for validation and 100 questions for testing. For the 1488 questions in the training dataset, we generated the equivalent question-question pair for each pair of questions in the same group. This is treated as the positive example. Then, we randomly sampled negative examples. Assume a question q in a group contained n equivalent questions, we randomly choose $2*n$ questions in different groups with q as negative sample set

for question q . And for each question in the training dataset, we get a negative sample set following the sampling strategy above and join all sets together as negative examples for supervised learning approaches.

Approaches

This section introduced the detailed methods we have explored and designed in this work. The first two methods, i.e. cosine similarity with Bag of word model (BOW-COS) and Word mover distance with word embedding (WE-WMD) are unsupervised methods and the other two methods, i.e. Supervised Siamese Recurrent Neural Network Model (SSRNN) and SNA, are supervised methods using deep neural networks. The BOW-COS is the baseline method and the SNA is an innovated neural network structure for calculating the similarity of two questions.

Cosine similarity with Bag of word model (BOW-COS)

BOW is a traditional text representation used in natural language processing and information retrieval. In this model, a question is represented as a vector of n dimension where n is total number of words in the corpus. Each dimension corresponds to a word, if the word appears in the question, the dimension is set to 1; otherwise it is set to 0. In addition, the Term Frequency-Inverse Document Frequency (TFIDF) weight is used to weight each word. This weight is a statistical measure used to evaluate how important a word t is to a document d in a corpus D contained N document and can be calculated as follows:

$$\text{tfidf}(t, d, D) = f_{t,d} \cdot \log(1 + \frac{N}{n_t}) \quad (1)$$

where $f_{t,d}$ means the frequency of word t in document d and n_t means the number of document containing word t in document collection D .

Cosine similarity is a traditional similarity algorithm for calculating the similarity score between two sentences or articles. The similarity score has a value of range 0-1. A score of zero means two questions are not similar at all, and a score of one means that they are completely the same. Assuming that there are two feature sets F_1 and F_2 that are generated from two different patient questions, the cosine similarity can be calculated as follows:

$$\text{COS}(F_1, F_2) = \frac{|F_1 \cap F_2|}{\sqrt{|F_1| \times |F_2|}} \quad (2)$$

Word mover distance with word embedding (WE-WMD)

The traditional BOW-COS calculates similarity of two questions based on their lexical overlap. However, the method failed to capture the semantic similarity of terms that are different. As the example mentioned previously, the BOW-COS method could not capture that the similarity between *diet* and *food*. This section introduces the word mover distance with word embedding (WE-WMD). This approach calculates word mover's distance which can leverage semantic relationships between words using the generated word embedding as aforementioned.

Word mover's distance (WMD) [14] is a distance function between two documents, which measures the minimum traveling distance from the embedded words of one document to another one. It allows each word embedding to be partially aligned to multiple word embedding of the other text. WMD achieved good performance in the document classification task. Specifically, given two document d and d' , WMD can be calculated by optimizing the following equation.

$$\min_{T \geq 0} \sum_{i,j=1}^n T_{ij} c(i, j) \text{ subject to: } \begin{cases} \sum_{j=1}^n T_{ij} = d_i \forall i \in \{1, \dots, n\} \\ \sum_{i=1}^n T_{ij} = d'_j \forall j \in \{1, \dots, n\} \end{cases} \quad (3)$$

where d_i means the weight of word i in document d and similarly, d'_j means the weight of word j in document d' , $c(i, j)$ means the distance between word i and word j and can be provided by Euclidean distance in the word2vec embedding space.

Supervised Siamese Recurrent Neural Network Model (SSRNN)

WE-WMD tries to solve the lexical gap problems by aligning similar words in word vector space. Recently, research interests have shifted toward learning word level similarity to the similarity of larger bodies of text such as sentences directly. Many supervised deep learning based models such as [8,9,15,16] use variants of neural network architectures to model question-question pair similarity.

In this research, we adopted one of the most popular architectures [15], Siamese RNN structure, to learn question similarity. The network structure consists of two share-weights LSTM networks each process one of the questions in a given question pair and a similarity measurement is used to calculate the similarity between the last states of the two LSTMs. Figure1 shows the network architecture in this research. Different from the original research which used Manhattan Distance and Mean Squared Error as loss function in sentence similarity task, we use similarity measurement as follows:

$$\text{Sim}(x, y) = (\text{COS}(x, y), \text{L2Euclid}(x, y), [x - y], x * y) \quad (4)$$

where (COS) and element-wise multiplication measure the distance of two vectors according to the angle between them, L2 Euclidean distance (L2Euclid) and element-wise absolute distance difference measure magnitude differences. In another perspective, COS and L2 Euclidean distance measure the sum distance and element-wise multiplication and absolute distance measure the element-wise distance in each dimension of the vectors. These four-measurement functions reflects similarity from different respect. Then we use a linear layer to combine these measurements together. In this research, L2 is set as 0.01. To train these model, we consider a classification task: given a pair of questions, the model will give a label of similar or not and use a log-softmax layer on top of the similarity measurement to output the label. After training, we use the probability of similar label to get question retrieval results.

We use stochastic gradient descent (SGD) to minimize the negative log likelihood loss function and the back propagation algorithm to compute the gradients. For the output layer, we employ dropout with a constraint on l2-norm of the weight vectors. The dropout rate is 0.2 and the l2 constraints is 3. Training is done through stochastic gradient descent over mini-batches with the size of 100 and Adadelta update rule [17] and the number of epochs is set to 200. The LSTM layer we use in this paper are bidirectional LSTM and the numbers of cell are all set as 20. The hidden state in the output linear layer is set as 50. We keep the word vectors static during training.

Supervised Neural Attention Model (SNA)

Recently, neural attention networks have demonstrated success in a wide range of tasks such as machine translation [18], image captioning [19], and sentence inference [20]. The idea is to allow the model to selectively focus on the most

task-relevant parts of input sequence, assign important weights to those parts, and join them into a single representation. In this research, we design a novel supervised neural attention model for similar question retrieval. Figure 2 shows network structure. Instead of calculating similarity between the last LSTM states, word by word attention mechanism is used to

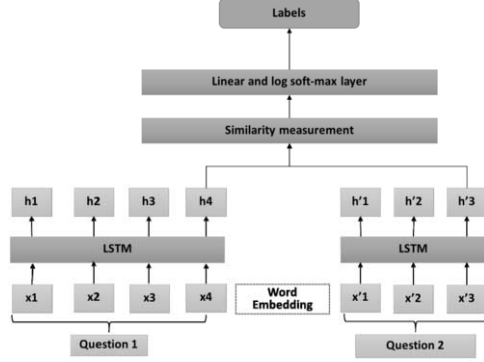


Figure 1 – Supervised Siamese Recurrent Neural Network structure used in this research

soft align words in the two questions. Specifically, the output state of the first LSTM attends the second LSTM's output vector and attention weights α_i and similarity measurement are calculated overall output vectors of Question 2 for each output state α_i in the Question 1. The word by word attention mechanism can be modeled as follows:

$$h_i^a = \left(\text{Sim}(\sum_{j=1}^m \alpha_{ij} h'_j, h_i), \sum_{j=1}^m \alpha_{ij} \text{Sim}(h'_j, h_i) \right) \quad (5)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_j \exp(e_{ij})} \quad (6)$$

$$e_{ij} = W^e \cdot \tanh(W^2 h'_j + W^1 h_i + W^s \text{Sim}(h'_j, h_i)) \quad (7)$$

where h_i^a is the similarity result between output state h_i of Question 1 and the output vectors of Question 2, all matrices W contain weights to be learned and the function Sim is same as Equation (4). Note that the similarity result h_i^a is a concatenation of two parts: 1) the similarity result between h_i and an aggregate of vector h' with the attention weights; 2) an aggregate of similarity result between h_i and each state h'_j . We also assume the soft align weight of α_{ij} is relevant to the similarity of h_i and h'_j .

After word by word attention, each output state of Question 1 has a similarity result with the Question 2 and then another LSTM is used to sequentially summarize the similarity result. We take the last state of the LSTM as similarity result and on top of the LSTM layer, we also use a linear layer and a log-softmax layer to output the label. The number of LSTM cell is also set to 20 and other setup is similar as SSRNN.

Table 3 shows the performance on the test dataset for the BOW-COS, WE-WMD, SSRNN with random initialized word embedding (SSRNN_RW), SSRNN with pre-trained word embedding, SNA with random initialized word embedding (SNA_RW), and SNA with pre-trained word embedding. Besides the supervised RNN models, we also compared our models with another state of art deep learning model RCNN

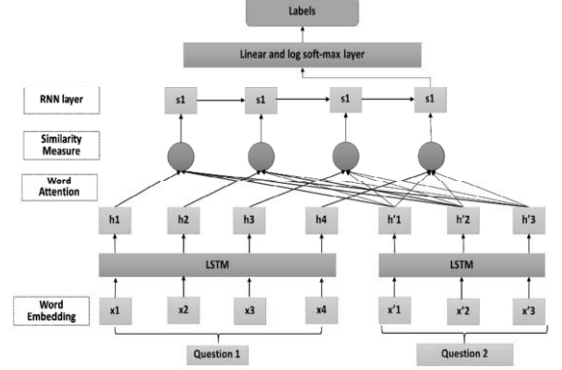


Figure 2 – Supervised Neural Attention Model

Results

Table 3 – Performances of the four approaches

Approach	MAP	MRR	P@1	P@5
BOW-COS	0.31	0.61	0.51	0.73
WE-WMD	0.36	0.65	0.54	0.79
SSRNN_RW	0.56	0.64	0.59	0.70
SSRNN	0.60	0.68	0.63	0.74
SNA_RW	0.57	0.71	0.64	0.82
SNA	0.69	0.80	0.77	0.84
RCNN	0.56	0.72	0.65	0.80

for question retrieval proposed in general domain [9], which designed a recurrent and convolutional model (gated convolution) to effectively map questions to their semantic representations. We retrieved the 100 test questions in 1488 training questions and 100 validating questions and evaluated the performance with three evaluation criteria: Mean Average Precision (MAP), Mean Reciprocal Rank (MRR), and Precision at K (P@K). From the result, we can see that approaches with word embedding (WE-WMD, SSRNN, and SNA) perform better than the similar approaches with random initialized word embedding (BOW-COS, SSRNN_RW and SNA_RW) and supervised deep learning models with pre-trained word vectors (SSRNN and SNA) perform better than unsupervised approach BOW-COS and WE-WMD. SNA with word by word attention mechanism performed best. Compared with the state of art RCNN approach, SNA also has superior performance.

Discussion

In this study, we investigated a novel supervised attention model to retrieve similar patient questions in Chinese. When only word features were used, approaches which utilized word embedding derived from another unlabeled corpus achieved better than the traditional BOW model. The performance improvement was from the semantic information automatically captured by word embedding. For example in Table 2, the semantically related words captured through word embedding may not have any relation in the bag of words representation space.

The supervised learning approaches, SSRNN and SNA performed better than the other unsupervised approaches. This may be because the supervised deep learning approaches can learn important features from the corpus automatically. But

SSRNN compares the last state between each LSTM and only captures a global-level representation of each questions. SNA compares each word from Question 1 with Question 2 through word by word attention mechanisms and can leverage details information of the two questions; SNA had the best performance.

Table 4 – An example for comparison between four approaches

Approach	我是冠心病患者，阿托伐他汀有效果吗 (I am a patient with coronary heart disease. Does Atorvastatin have an effect on this?)
BOW-COS	1 阿托伐他汀是保健品吗 (Is Atorvastatin a health care product?) 2 20mg阿托伐他汀是有两种包装吗 (Does 20 mg Atorvastatin have two kinds of packages?)
WE-WMD	1 阿托伐他汀是治疗什么疾病的药物？ (Which disease is Atorvastatin used to treat?) 2 两种阿托伐他汀的药效有区别吗？ (Are there any difference between the pharmacodynamics of the two kinds of Atorvastatin)
SSRNN	1阿托伐他汀对血脂指标的作用？ (Does Atorvastatin have an effect on blood lipids?) 2阿托伐他汀是治疗血脂高的吗？ (Can Atorvastatin be used for treating hyperlipemia?)
SNA	1阿托伐他汀是治疗什么疾病的药物？ (Which disease is Atorvastatin used to treat?) 2阿托伐他汀的主要作用？ (What is the main usage of Atorvastatin?)

In Table 4, an example for comparing between the four approaches is shown. The top two questions retrieved by BOW-COS are both irrelevant with the query question. In WE-WMD, as the word 效果 (*effect*) and the word 药效 (*pharmacodynamics*) is semantically-related in word embedding, the irrelevant question 两种阿托伐他汀的药效有区别吗？ (*Are there any difference between the pharmacodynamics of the two kinds of Atorvastatin*) is retrieved by the system. In SSRNN, the two retrieved questions are about Atorvastatin's effect on blood lipids. This is because SSRNN focuses on the global representation such as Atorvastatin's effect but ignores the details such as blood lipids vs. coronary heart disease. However, the SNA model considers the difference of details through word by word attention.

As current research about question retrieval in the health care domain still focuses on feature engineering such as medical concept annotation by UMLS or sentence parser, but obtaining such features could be time consuming and sometimes not feasible in other languages such as Chinese. However, using a deep learning based approaches requires minimal effort on feature engineering.

Although the deep learning methods in this work show superiority in performance, human annotation dataset is still need for supervised learning. And the difference between train and test data will cause performance reduction. In the future, we plan to investigate different approaches to further reduce human effort such as transfer learning and integrate medical knowledge with the deep learning network structure.

Conclusion

In this study, we investigated a deep learning-based model for similar patient question retrieval in Chinese. Our results showed that our SNA model performed best. Further analysis showed semantic information automatically captured by word

embedding is helpful and word by word attention mechanisms which can model detailed differences can further improve the performance.

References

- [1] S. Fox, M. Duggan, Health online 2013. <http://www.webcitation.org/6eAaGTaU> Pew Research center, 2013
- [2] J. Jiwoon, W.B. Croff, J.H. Lee, Finding Similar Questions in Large Question and answer archives. In: *Conference on Information and Knowledge Management* 2005.
- [3] X. Xue, J. Jeon, W.B. Croft, retrieval models for question and answer archives. In: *Proceedings of the 31st Annual International ACM SIGIR Conference*, 2008.
- [4] G. Zhou, L. Cai, J. Zhao, K. Liu, Phrase-based translation model for question retrieval in community question answer archives. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.
- [5] L. Cai, G. Zhou, K. Liu, J. Zhao, Learning the latent topics for question retrieval in community QA. In: *Proceedings of the 5th International Joint Conference Natural Language Processing*, 2011.
- [6] Z. Ji, F. Xu, B. Wang, B. He, Question-answer topic model for question retrieval in community question answering. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2012.
- [7] K. Zhang, W. Wu, H. Wu, Z. Li, M. Zhou, Question retrieval with high quality answers in community question answering. In: *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management*, 2014.
- [8] A. Das, H. Yenala, M. Chinnakotla, M. Shrivastava, Together we stand: siamese networks for similar question retrieval. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* 2016.
- [9] Tao Lei, Hrishikesh Joshi, Regina Barzilay, Tommi Jaakola, Katerina Tymoshenko, Alessandro Moschitti and Luis Marquez. Semi-supervised Question Retrieval with Gated Convolutions. In *Proceedings of NAACL-HLT*, 2016.
- [10] J. Luo, G.Q. Zhang, S. Wentz, L. Cui, R. Xu, SimQ: real-time retrieval of similar consumer health questions. *J Med Internet Res* **17** (2015):e43.
- [11] P. Wongchaisuwat, D. Klabjan, S.R. Jonnalagadda, A semi-supervised learning approach to enhance health care community-based question answering: a case study in alcoholism. *JMIR Med Inform* **4** (2016):e24.
- [12] Z. Shaojian, T. Kang, X. Zhang, D. Wen, N. Elhadad, J. Lei. Speculation detection for Chinese clinical notes: Impacts of word segmentation and embedding models. *J Biomed Inform* **60** (2016):334-341.
- [13] T. Mikolov, K. Chen, G. Corrado, J. Dean, efficient estimation of word representations in vector space. In: *arXiv:1301.3781*, 2013.
- [14] M.J. Kusner, Y. Sun, N.I. Kolkin, K.Q. Weinberger, From word embeddings to document distances. In: *Proceedings of the 32nd International Conference on Machine Learning* **2015** (2015): 957-966.
- [15] J. Mueller, A. Thyagarajan, Siamese recurrent architectures for learning sentence similarity. In: *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [16] X. Qiu, Xipeng, X. Huang, Convolutional neural tensor network architecture for community-based question answering. In: *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 2015.
- [17] M. D. Zeiler. Adadelata: an adaptive learning rate method, *CoRR*, 2012.
- [18] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate. In: *arXiv preprint arXiv:1409.0473*, 2014.
- [19] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, Y. Bengio, Show, attend and tell: Neural image caption generation with visual attention. In: *Proceedings of The 32nd International Conference on Machine Learning* (2015):2048-2057.
- [20] T. Rocktäschel, E. Grefenstette, K.M. Hermann, T. Kočiský, P. Blunsom, Reasoning about entailment with neural attention. In: *arXiv preprint arXiv:1509.06664*, 2015.

Address for correspondence

Corresponding author name: Guoyu Tang

Email: bjtanggy@cn.ibm.com (preferred method of contact)

Mailing address: Building 19 Zhongguancun Software Park,

8 Dongbeiwang WestRoad, Haidian District,

Beijing, P.R.China.100193

Phone: 86-10-58748158

Automating the Identification of Patient Safety Incident Reports Using Multi-Label Classification

Ying Wang^a, Enrico Coiera^a, William Runciman^{b, c}, Farah Magrabi^a

^a Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Australia

^b Centre for Population Health Research, School of Health Sciences, University of South Australia, Australia

^c Australian Patient Safety Foundation, Adelaide, Australia

Abstract

Automated identification provides an efficient way to categorize patient safety incidents. Previous studies have focused on identifying single incident types relating to a specific patient safety problem, e.g., clinical handover. In reality, there are multiple types of incidents reflecting the breadth of patient safety problems and a single report may describe multiple problems, i.e., it can be assigned multiple type labels. This study evaluated the ability of multi-label classification methods to identify multiple incident types in single reports. Three multi-label methods were evaluated: binary relevance, classifier chains and ensemble of classifier chains. We found that an ensemble of classifier chains was the most effective method using binary Support Vector Machines with radial basis function kernel and bag-of-words feature extraction, performing equally well on balanced and stratified datasets, (F-score: 73.7% vs. 74.7%). Classifiers were able to identify six common incident types: falls, medications, pressure injury, aggression, documentation problems and others.

Keywords:

Machine Learning; Risk Management; Patient Safety

Introduction

Approximately 10% of admissions to acute-care hospitals are associated with unnecessary harm to patients [1, 2]. Events that could have resulted, or did result in unnecessary harm are called *patient safety incidents* [3]. The reporting and rapid analysis of patient safety incidents is important to prevent similar events from occurring in the future [4]. However, the volume of incident reports has dramatically increased over the last 20 years with wide implementation of incident monitoring systems [5]. Retrospective review of these incident reports by human experts is highly resource intensive and can no longer keep up with the growing volume of incidents being reported. To facilitate timely analysis, many incident monitoring systems now ask reporters to assign incident types so that reports of a specific type can be easily grouped for detailed classification by experts. However, incidents are reported by many different groups of health professionals who generally have limited expertise in classification [3, 6, 7].

One way of improving the efficiency of identifying incidents of a specific type is to automatically classify reports using text classification techniques. We have previously shown the feasibility of using statistical text classification to identify reports about three types of incidents: patient identification [8], clinical handover [8] and health information technology [9] using binary classifiers based on Naïve Bayes, logistic

regression and Support Vector Machines (SVM). We have subsequently shown that extreme-risk events could be identified using a similar approach [10]. However, these studies have focused on distinguishing specific incident types from all others, e.g., patient identification. In reality, there are multiple types of incidents reflecting the breadth of problems in patient safety. Moreover, a single report can describe problems in more than one patient safety area, i.e., it can be assigned to multiple incident types [3, 11]. For example, "Episode label A for patient X was placed incorrectly onto the specimen that belongs to patient Y," describes an error in patient identification that also relates to documentation.

The use of automated methods to identify multiple incident types remains largely unexplored. Some studies have sought to apply topic modelling [12, 13]. However, the mapping between topics and incident types is not straightforward. In this paper, we evaluated the feasibility of using multi-label classification to automate the identification of two labels or two incident types per report. Such multi-label problems are typically decomposed into one or more binary problems to simplify decision boundaries [14, 15]. In this study, classifiers were trained, validated and tested on balanced datasets. We then examined generalizability by applying the classifiers to imbalanced, i.e., stratified datasets which represented the real-world distribution of incidents.

Methods

Multi-label classification was decomposed into a number of binary classification problems, one for each label. We sought to evaluate the performance of three decomposition methods: 1. *Binary Relevance* (BR) is a widely used decomposition method which is theoretically simple and intuitive. For a given set of labels L , BR learns L binary classifiers (e.g., SVM). Each classifier is trained independently for a specific label against the rest of labels (one-versus-rest scheme) [15]. However, it does not preserve dependencies between labels. 2. *Classifier Chains* (CC) is another decomposition method which is based on the BR but considers dependencies between labels [16]. CC learns L binary classifiers like the BR method. However, these binary classifiers are linked in a chain through a feature space where additional features with binary values indicate which other labels are assigned to a report. Testing begins with the first classifier and processes to the L th classifier by passing label information between classifiers through this feature space. Hence, the inter-label dependency is preserved. The performance of CC is very sensitive to the choice of label order in the chain, as the label dependency is demonstrated by label order in the feature space.

3. Ensemble of Classifier Chains (ECC) were introduced to reduce the influence of label order and improve classification performance [16]. Multiple CCs can be trained using different random chain ordering (determined by the order of labels) on a random subset of data. The final decision is the average of the multi-label predictions of CC. In general, the performance of ECC tends to improve and converge by combining more CCs with diverse label structures. However with increasing ensemble sizes, ECC becomes unnecessarily large, and imposes an extra computational cost. Although there are several variants of CC which consider complex structures for label dependence and use different random search methods [17], we took a more practical solution by using a subset of CCs in the ensemble, which also improved classification performance.

Database

We used 6,000 randomly selected reports from 137,522 submitted to the Advanced Incident Management System (AIMS) across an Australian state between January and December 2011 [4]. Ethical approval was obtained from university committees as well as a committee governing the hospital and state datasets. Reports were de-identified and then labelled by three experts in the classification of patient safety incident reports. Experts provided a primary label and where applicable a secondary label was also given. These labels were used as a "gold standard" for classifier training and testing. Only descriptive narratives in reports were retained for experiments including incident description, patient outcome, actions taken, prevention steps, investigation findings and results. All system-specific codes, punctuation and non-alphanumeric characters were removed and text was converted to lower case.

The distribution of incident types from AIMS is imbalanced i.e. "stratified", and their real-world ratio is shown in Table 1.

Table 1: The composition of incident types in balanced and stratified datasets. N1 is the number of reports based on primary label. N2 is the number of reports by considering both labels.

Incident type	Training: balanced datasets		Testing: stratified datasets		
	N1	N2	N1	N2	%
Falls	260	261	90	91	20
Medications	260	304	68	74	15
Pressure injury	260	264	37	38	8
Aggression	260	271	49	57	11
Documentation	260	589	26	67	6
Blood product	260	273	5	6	1
Patient identification	260	337	7	8	2
Infection	260	274	6	6	1
Clinical handover	260	301	7	8	2
Deteriorating patient	260	264	1	2	<1
Others	260	689	148	173	33
Total	2860	3827	444	530	

To build reliable classifiers which capture characteristics of rare incident types, for training we used balanced AIMS datasets where each of the 10 types were evenly distributed (Table 1). Applicability to real-world conditions was then examined by testing on stratified data. The distribution of balanced and stratified datasets was based on primary labels.

Experimental workflow

Figure 1 shows an overview of our approach. Datasets were first decoded into 11 subsets according to one-vs-rest ensemble schemes [18, 25]. For the BR, 11 base classifiers were trained for each two-class subset involving feature extraction, classifier training and cross validation. A threshold-based decision-making scheme was applied to identify testing reports by combining the predicted probabilities from all base classifiers. With CC, the feature space was extended to represent label connections using binary values that indicated label co-occurrence [18, 19]. With ECC, incident types were randomly reordered and several ensembles of CC were generated. The hard decision for each report was made based on the probability values provided by all CCs.

1. **Feature extraction:** To provide informative features for classification, removal of stop words and short words with fewer than two characters, stemming and lemmatization were applied to reports [18]. The bag-of-words model, commonly used in document classification, was adopted to extract features [19]. Irrespective of grammar, incident narratives were represented as an unordered collection of words and unique words were used as features. The bag of words was then transformed into a numeric representation interpretable by classifiers. A binary count, transforming the bag of words representation into 1 or 0 corresponding to word occurrences, was used, as it was the most effective feature representation along with the SVM classifier in practice [9, 10].

2. **Feature space extension for CC and ECC:** The feature space of each link in the classifier chain was extended with the binary label associations of all links in the training procedure [16]. The order of the chain itself affects accuracy.

3. **Base binary classifier training and validation:** For base classifiers, we chose discriminative classifier of SVM with radial-basis function kernel, as the SVM-based methods perform better for smaller datasets with a large feature space [20, 21]. Especially in text classification, documents are typically represented as a bag-of-words, where each feature captures crucial information but in a very sparse format [20]. Furthermore, in our previous incident classification work, SVM outperformed other binary classifiers, such as a logistic regression model [21].

To train the base classifiers, a ten-fold repeated random sub-sampling, cross-validation method was used to assign reports

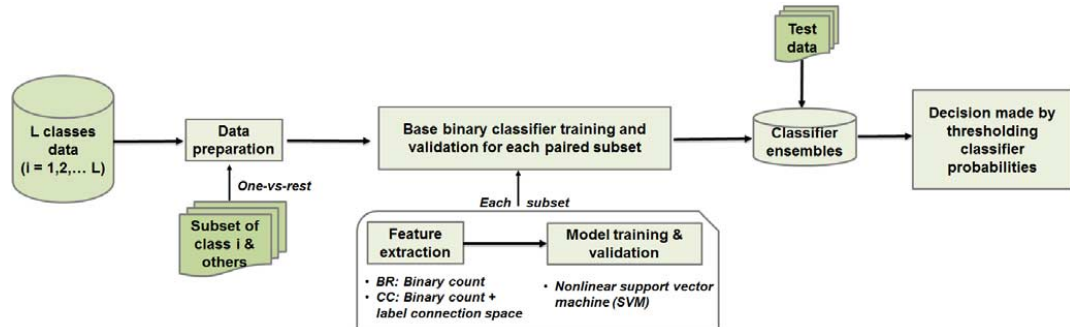


Figure 1: Experimental workflow to train and evaluate methods of binary relevance (BR) and classifier chains (CC) to identify reports. Ensemble classifier chains (ECC) is a repetition of CC using randomly ordered labels.

to training (80%), validation (10%) and testing (10%) sets. Given that, with random assignment, a testing report for one base binary classifier might be used in the training set of another base classifier. For example, a report about a fall incident which was used for training a base classifier (fall vs. medications), may also be assigned to the test set of another base classifier (fall vs. blood product). To avoid potential overlaps between training and testing sets, we first randomly selected 10% of reports for each type and set them aside for testing. Then we created the folds using repeated random subsampling for training and validation sets. Classifiers which outperformed others by achieving higher classification accuracy were adopted for testing.

4. Probability threshold based decision-making scheme: For ECC, the hard decision was made by averaging multiple predictions from individual CC. Two labels were predicted if the averaged classification probabilities exceeded a predefined threshold. If the threshold is reached only for one label then the report was assigned to a single incident type.

5. Performance evaluation: Evaluation metrics for multi-label classification performance are inherently different from those used in multi-class or binary classification, due to the additional degrees of freedom that the multi-label setting introduces. Overall performance was examined using two types of evaluation measures, example-based and label-based measures [20]. Example-based measures are based on the average difference of the true and the predicted sets of labels over all testing reports. On the other hand, label-based measures evaluate classification performance separately for each incident type and then average the performance over all types. We used six example-based measures including [20]:

1. *Hamming loss* evaluates how many times an incident is misclassified, i.e. label not belonging to the report is predicted or a label belonging to the report is not predicted. This is a loss function, so the optimal value is zero:

$$\text{Hamming loss} = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=1}^L \text{xor}(T_{ij}, P_{ij})}{L} \quad (1)$$

Where L is the total number of incident types and N is the number of testing reports. X_{or} means the symmetric difference between two sets. T_{ij} denotes the set of true labels and P_{ij} denotes the set of predicted labels.

2. *Accuracy* is micro-averaged across all reports and is defined as the number of correct labels divided by the union of predicted and true labels.

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{|T_i \cap P_i|}{|T_i \cup P_i|} \quad (2)$$

Note that any predicted label matches the real labels, count 1; so the size of correct labels for a report could be 0, 1, and 2 if there are two real labels. However, the number of correct labels in this way does not consider label orders.

3. *Exact match score* (0/1 loss) is a very strict evaluation measure as it requires the predicted set of labels to be an exact match of the true set of labels.

4. *Precision* is the correct labels divided by the number of number of predicted labels.

$$\text{Precision} = \frac{1}{N} \sum_{i=1}^N \frac{|T_i \cap P_i|}{|P_i|} \quad (3)$$

5. *Recall* is defined as the correct labels divided by the number of number of true labels.

$$\text{Recall} = \frac{1}{N} \sum_{i=1}^N \frac{|T_i \cap P_i|}{|T_i|} \quad (4)$$

6. *F-score* is the harmonic mean between precision and recall.

$$\text{Fscore} = (1 + \beta^2) \frac{\text{precision} * \text{recall}}{(\beta^2 * \text{precision}) + \text{recall}} \quad (5)$$

Where β enables the F-score to favor either precision or recall. Precision and recall are given equal weight by setting β to 1.

We also used six label-based measures, which are defined below. Given two labels per report, the classification performance of individual types was evaluated based on OR logic when matching the predicted and true set of labels per report.

1. *Macro-precision* is defined as precision averaged across all labels:

$$\text{Macro-precision} = \frac{1}{L} \sum_{j=1}^L \frac{|tp_j|}{|tp_j + fp_j|} \quad (6)$$

Where tp_j and fp_j are the number of true positives and false positives for the label j considered as a binary class.

2. *Macro-recall* is defined as recall averaged across all labels:

$$\text{Macro-recall} = \frac{1}{L} \sum_{j=1}^L \frac{|tp_j|}{|tp_j + fn_j|} \quad (7)$$

Where fn_j is the number of false negatives for the label j which is considered as a binary class.

3. *Macro-F-score* is the harmonic mean between precision and recall where the average is calculated per label and then averaged across all labels.

$$\text{Macro-F-score} = \frac{1}{L} \sum_{j=1}^L \frac{2 * \text{precision}_j * \text{recall}_j}{\text{precision}_j + \text{recall}_j} \quad (8)$$

Micro-precision and micro-recall are defined as general definitions of precision and recall but the number of true positives, false positives and false negatives are averaged over all the labels, respectively. Micro-F-score is the harmonic mean of micro-precision and micro-recall, defined as below.

$$\text{Micro-precision} = \frac{\sum_{j=1}^L tp_j}{\sum_{j=1}^L tp_j + \sum_{j=1}^L fp_j} \quad (9)$$

$$\text{Micro-recall} = \frac{\sum_{j=1}^L tp_j}{\sum_{j=1}^L tp_j + \sum_{j=1}^L fn_j} \quad (10)$$

$$\text{Micro-score} = \frac{2 * \text{micro}_{\text{precision}} * \text{micro}_{\text{recall}}}{\text{micro}_{\text{precision}} + \text{micro}_{\text{recall}}} \quad (11)$$

Results

We examined the performance of three multi-label classification methods on balanced and stratified datasets. Using the most effective classifiers, we then evaluated performance when identifying different incident types.

ECC performance with increasing ensemble sizes

Using SVM RBF with binary count we examined the ECCs using different ensemble sizes ranging from three to forty. We found that six classifier chains obtained much smaller ensembles while achieving better or at least comparable performance on balanced and stratified datasets (Table 2, columns 3-7).

Performance of BR, CC and ECC

We compared BR, CC and ECC with six CCs (Table 2). Performance improved from BR to CC to ECC with both balanced and stratified datasets. Considering exact matching between the predicted and true sets of incident types, ECC (balanced/stratified: 39.9%, 44.4%) outperformed BR (35.7%, 39.0%) and CC (36.7%, 37.8%), indicating that better label dependency was maintained by ECC. Hamming loss value also showed that ECC was more efficient than BR. Example-based F-scores and label-based Macro-F-scores and Micro-F-scores increased slightly from BR to ECC with both testing datasets, for instance, F-scores of BR: 73.7%, CC: 74.2% and ECC: 74.7% on the stratified datasets. The most effective ECC performed equally well on both balanced and stratified datasets, considering example-based measures and Micro-F-score, indicating that this approach is generalizable. However, it obtained a relatively worse Macro-F-score of 59.2% on stratified datasets (balanced: 73.7%).

Table 2: Classifier performance of BR, CC, and ECC with different ensemble sizes on balanced and stratified AIMS datasets, where hamming loss, accuracy, recall, precision, F-score, extract match score, Macro/Micro-averaged measures were considered. Base binary classifier was SVM RBF with binary count representation of bag-of-word features.

Measures (%)	BR		CC		ECC					
					3CC	6CC	11CC	20CC	40CC	
Hamming Loss	8.4	8.3	8.1	7.8	7.8	7.9	7.9	7.7		
	8.0	8.2	7.6	7.2	7.5	7.5	7.5	7.5		
Accuracy	62.6	63.5	64.4	64.4	65.1	64.2	65.3			
	65.4	65.6	67.1	68.0	66.8	66.7	66.6			
Exact match score	35.7	36.7	39.9	39.9	40.6	40.2	40.6			
	39.0	37.8	43.0	44.4	44.4	43.5	43.5			
Precision	67.7	68.4	70.3	70.6	71.3	70.5	71.9			
	68.9	69.3	71.1	72.9	71.7	71.6	71.6			
Recall	80.1	80.4	77.1	77.1	77.8	76.7	77.8			
	79.1	80.0	76.6	76.6	75.5	75.9	75.6			
F-score	73.3	73.9	73.5	73.7	74.4	73.5	74.7			
	73.6	74.2	73.7	74.7	73.6	73.7	74.4			
Macro-precision	67.2	66.2	69.0	69.7	69.9	70.0	70.0			
	50.0	49.0	51.0	52.4	52.3	52.2	51.9			
Macro-recall	80.7	80.0	78.9	79.0	79.6	79.2	79.5			
	80.4	79.7	77.5	77.4	77.6	77.4	77.3			
Macro-F-score	72.9	72.2	73.2	73.7	74.0	73.9	74.0			
	57.6	56.7	57.9	59.2	59.0	58.9	58.6			
Micro-precision	63.3	63.3	66.1	67.1	67.1	67.3	67.4			
	62.8	62.3	66.0	67.1	67.0	66.9	66.6			
Micro-recall	73.4	73.4	71.9	71.9	72.6	71.9	72.3			
	71.5	72.2	71.2	70.6	70.7	70.5	70.4			
Micro-F-score	68.0	68.0	68.9	69.4	69.7	69.5	69.7			
	66.9	66.9	68.5	68.8	68.8	68.7	68.4			

* measures from stratified datasets are shaded.

Performance on identifying individual incident types

The most effective ECC based on six CCs achieved high F-scores above 81% on balanced datasets when identifying falls, pressure injury, infection and deteriorating patients (Table 3). This ECC trained on balanced datasets was very robust in identifying six types of incidents (falls, medication, pressure injury, aggression, documents and others), achieving similar F-scores on stratified datasets. For blood products, patient identification, infection, clinical handover and deteriorating patients, recall was similar on both balanced and stratified datasets, but precision was lower on stratified datasets.

Table 3: Individual incident type classification performance (recall, precision and F-score) on balanced (B) and stratified (S) datasets using 6 CC of binary SVM RBF with binary count

Incident type	Recall (%)		Precision (%)		F-score (%)	
	B	S	B	S	B	S
Falls	96.3	96.8	81.3	83.5	88.1	89.7
Medication	73.8	77.9	62.0	74.7	67.4	76.3
Pressure injury	90.0	90.5	93.1	80.9	91.5	85.4
Aggression	87.1	80.3	64.3	60.6	74.0	69.1
Documentation	62.2	62.6	62.2	59.8	62.2	61.2
Blood products	81.8	85.7	62.8	30.0	71.1	44.4
Patient identification	73.9	72.7	59.6	22.2	66.0	34.0
Infection	83.3	85.7	78.9	31.6	81.1	46.2
Clinical handover	76.9	72.7	52.6	20.5	62.5	32.0
Deteriorating patient	96.3	66.7	83.9	33.3	89.7	44.4
Others	53.9	59.6	67.6	78.9	60.0	67.9

Discussion

We evaluated three multi-label text classification methods using binary classifier ensemble on both balanced and stratified datasets from a state-wide incident reporting system. The most effective classifiers performed equally well in identifying reports from six common types including falls, pressure injury, aggression, documents and others. These types made up 93% of all reports (Table 1). Even so, it should be emphasized that automated identification of incident reports is not intended as a replacement for expert review but as a first step in grouping incidents and identifying clusters when human resources are lacking [22].

Comparison between BR, CC and ECC

Although the ECC achieved the best performance, the performance difference between BR, CC and ECC was relatively small (Table 2). Compared to CC and ECC, BR is more computationally efficient as it requires less dimensionalities of features and no classifier chain for training. However the BR completely ignores the possible correlations among labels, so the binary classifiers make decisions independently from each other. On the other hand, it makes BR very flexible in practice, e.g., evolving more incident types, because it can add labels without affecting the rest of classifiers. Thus, BR is more suitable for problems with only a small number of labels associating with each other.

The CC method can be seen as a direct extension of the BR, capable of exploiting label dependencies. Similar to BR, CC involves training a group of L binary classifiers. Nevertheless, instead of being kept isolated from each other, these L classifiers are linked in a chain structure, which allows each one to pass their predictions to the other binary classifiers connected in the chain. Compared to BR, CC did not improve much of classification performance in this study. BR was even slightly better than CC according to hamming loss and exact matching score on the stratified datasets. This might be due to the original text feature space being huge and the label relationship in such a sparse space not being captured well by classifiers with limited training datasets.

ECC keeps relationship of patient safety types in reports by building multiple CCs using random label orders. Obviously, the larger ensemble size leads to higher computational cost. We tried to reduce the ensemble size whilst keeping or improving the performance. In training, we built several ECCs with ensemble sizes ranging from three to forty. We found that the classification performance became stable when the size of CC was greater than six (Table 2). Thus the most effective ECC used six CCs was used for testing.

Identification of common and rare incident types

An important finding of this study is that the most effective ECC trained on balanced datasets appears generalizable to unseen datasets. The ECC achieved similar F-scores on stratified datasets in identifying the common incident types including falls, medication, pressure injury, aggression, documentation and others (4-33% Table 1, 93% of reported incidents in total). Especially for falls and pressure injury, the ECC achieved high F-scores (above 85.4%). However the ECC tended to be weaker when identifying rare types in stratified datasets such as patient identification, infection, clinical handover and deteriorating patient (<2% Table 1, 7% of all reported incidents). This explains why similar example-based measures were achieved on balanced and stratified datasets because the common types dominate the performance measures. On the other hand, the label-based measures showed relatively worse performance on stratified datasets, as the averaged measures were based on individual performance

of each type. Given the imbalanced nature of incident distribution, automated identification can reduce the effort spent in identifying common types and provide small volumes of like incident reports for further investigation by experts.

Limitations and future work

There are several limitations. First, we used datasets from one Australian state. Our classifiers may not be generalizable to other regions with different linguistic styles and terminology. Secondly, we use the balanced dataset for classifier training because a limited number of incidents had been reported over a 12-month period. Given the imbalanced nature of multiple incident types, a stratified training set may work better in real-world conditions. To improve the identification of rare classes, one solution might be to review rare classes flagged by classifiers, which is practical because overall volumes in real-world datasets will be low. Another possible way is to use rule-based methods that involve expert knowledge and incorporate specific criteria for identifying incidents.

Conclusions

The use of text-based ECC is a feasible approach for automatically identifying multiple incident types. Evaluation of BR, CC, and ECC using binary classifiers of SVM RBF with binary count feature extraction, showed that the most effective combination was the six ECC of binary SVMs. Despite the limitations listed above, automated identification can provide a more efficient way to categorize the common incident reports, so that human resources can be redirected to detailed classification, allowing remedial actions to be triggered more quickly to respond to emerging safety issues.

Acknowledgements

We thank B Shumack, K Pappas and D Arachi for assisting with the data extraction and, A Deakin, A Agers and S Suffolk for their assistance with labeling the incident reports.

References

- [1] N. Rafter, A. Hickey, S. Condell, R. Conroy, P. O'Connor, D. Vaughan, and D. Williams, Adverse events in healthcare: learning from mistakes, *QJM* **108** (2015), 273-277.
- [2] W.B. Runciman, R.K. Webb, S.C. Helps, E.J. Thomas, E.J. Sexton, D.M. Studdert, and T.A. Brennan, A comparison of iatrogenic injury studies in Australia and the USA. II: Reviewer behaviour and quality of care, *Int J Qual Health Care* **12** (2000), 379-388.
- [3] G. World Alliance For Patient Safety Drafting, H. Sherman, G. Castro, M. Fletcher, S. World Alliance for Patient, M. Hatlie, P. Hibbert, R. Jakob, R. Koss, P. Lewalle, J. Loeb, T. Perneger, W. Runciman, R. Thomson, T. Van Der Schaaf, and M. Virtanen, Towards an International Classification for Patient Safety: the conceptual framework, *Int J Qual Health Care* **21** (2009), 2-8.
- [4] W.B. Runciman, J.A. Williamson, A. Deakin, K.A. Benveniste, K. Bannon, and P.D. Hibbert, An integrated framework for safety, quality and risk management: an information and incident management system based on a universal patient safety classification, *Qual Saf Health Care* **15 Suppl 1** (2006), i82-90.
- [5] NRLS Quarterly Data Workbook up to September 2015, in, 2016, p. NRLS Quarterly Data Workbook updates analysis of the NRLS patient safety incidents reported by NHS organisation in England and Wales to the National Reporting and Learning System up to September 2015.
- [6] J.F. Travaglia, M.T. Westbrook, and J. Braithwaite, Implementation of a patient safety incident management system as viewed by doctors, nurses and allied health professionals, *Health (London)* **13** (2009), 277-296.
- [7] J. Westbrook, L. Li, E. Lehnborn, M. Baysari, J. Braithwaite, R. Burke, C. Conn, and R. Day, What are incident reports telling us? A comparative study at two Australian hospitals of medication errors identified at audit, detected by staff and reported to an incident system, *International Journal for Quality in Health Care* **27** (2015), 1-9.
- [8] M.S. Ong, F. Magrabi, and E. Coiera, Automated categorisation of clinical incident reports using statistical text classification, *Qual Saf Health Care* **19** (2010), e55.
- [9] K.E.K. Chai, S. Anthony, E. Coiera, and F. Magrabi, Using statistical text classification to identify health information technology incidents, *Journal of the American Medical Informatics Association* **20** (2013), 980-985.
- [10] M.S. Ong, F. Magrabi, and E. Coiera, Automated identification of extreme-risk events in clinical incident reports, *J Am Med Inform Assoc* **19** (2012), e110-118.
- [11] National Safety and Quality Health Service Standards Report, *Australia Commission on Safety and Quality in Health Care* (2012).
- [12] A. Fong, A. Hettinger, and R. Ratwani, Exploring methods for identifying related patient safety events using structured and unstructured data, *J Biomed Inform* **58** (2015), 89-95.
- [13] A. Fong and R. Ratwani, An Evaluation of Patient Safety Event Report Categories Using Unsupervised Topic Modeling, *Methods Inf Med* **54** (2015), 338-345.
- [14] J. Read, B. Pfahringer, G. Holmes, and E. Frank, Classifier chains for multi-label classification, *Machine Learning* **85** (2011), 333-359.
- [15] G. Tsoumakas and I. Katakis, Multi-label classification: An overview, *Dept. of Informatics, Aristotle University of Thessaloniki, Greece* (2006).
- [16] J. Read, B. Pfahringer, G. Holmes, and E. Frank, Classifier Chains for Multi-label Classification, *Machine Learning and Knowledge Discovery in Databases, Pt II* **5782** (2009), 254-269.
- [17] E.C. Goncalves, A. Plastino, and A.A. Freitas, Simpler is Better: a Novel Genetic Algorithm to Induce Compact Multi-label Chain Classifiers, *Proceedings of the 2015 Genetic and Evolutionary Computation Conference* (2015), 559-566.
- [18] T. Korenius, J. Laurikkala, K. Järvelin, and M. Juhola, Stemming and lemmatization in the clustering of Finnish text documents, in: *Proceedings of the thirteenth ACM international conference on Information and knowledge management*, ACM, 2004, pp. 625-633.
- [19] J. Sivic and A. Zisserman, Efficient Visual Search of Videos Cast as Text Retrieval, *Ieee Transactions on Pattern Analysis and Machine Intelligence* **31** (2009), 591-606.
- [20] G. Madjarov, D. Kocev, D. Gjorgjevikj, and S. Dzeroski, An extensive experimental comparison of methods for multi-label learning, *Pattern Recognition* **45** (2012), 3084-3104.
- [21] C. Cortes and V. Vapnik, Support-Vector Networks, *Machine Learning* **20** (1995), 273-297.
- [22] C.P. Friedman, A "fundamental theorem" of biomedical informatics, *J Am Med Inform Assoc* **16** (2009), 169-170.

Address for correspondence

Ying Wang, Email: ying.wang@mq.edu.au.

Mining Adverse Events of Dietary Supplements from Product Labels by Topic Modeling

Yefeng Wang^a, Divya R. Gunashekar^b, Terrence J. Adam^{a,c}, Rui Zhang^{a,d}

^a Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA,

^b School of Public Health, University of Minnesota, Minneapolis, MN, USA,

^c College of Pharmacy, University of Minnesota, Minneapolis, MN, USA,

^d Department of Surgery, University of Minnesota, Minneapolis, MN, USA

Abstract

The adverse events of the dietary supplements should be subject to scrutiny due to their growing clinical application and consumption among U.S. adults. An effective method for mining and grouping the adverse events of the dietary supplements is to evaluate product labeling for the rapidly increasing number of new products available in the market. In this study, the adverse events information was extracted from the product labels stored in the Dietary Supplement Label Database (DSLDD) and analyzed by topic modeling techniques, specifically Latent Dirichlet Allocation (LDA). Among the 50 topics generated by LDA, eight topics were manually evaluated, with topic relatedness ranging from 58.8% to 100% on the product level, and 57.1% to 100% on the ingredient level. Five out of these eight topics were coherent groupings of the dietary supplements based on their adverse events. The results demonstrated that LDA is able to group supplements with similar adverse events based on the dietary supplement labels. Such information can be potentially used by consumers to more safely use dietary supplements.

Keywords: Dietary Supplements; Natural Language Processing; Pharmacovigilance

Introduction

According to a recent cross-sectional study using the data from the National Health and Nutrition Examination Survey (NHANES) over the time period from 199-2012, 52% of U.S. adults use dietary supplements [1]. Americans spend more than \$28 billion annually on dietary supplements in the belief that they are able to improve health and are generally harmless in their side effects [2]. However, a study conducted by the Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention (CDC) has shown that dietary supplements caused on average about 23,000 emergency room visits every year [3].

The distinct regulatory framework for dietary supplements is one of the major reasons why dietary supplement safety requires scrutiny. In 1994, the Dietary Supplement Health and Education Act (DSHEA) was passed. According to DSHEA, dietary supplements belong to a subcategory of food, and can be waived from the premarket approval for efficacy and safety testing that are required for new drugs and food additives [4, 5]. While the DSHEA did give the consumers a great variety of choices in dietary supplements, it also exposed consumers to considerable risks in terms of dietary supplement safety [6].

A major clinical concern is that the adverse events of the active ingredients in the supplements may be unknown. For example, the sexual enhancement supplement Zotrex, which had

been recalled by FDA in 2011, contained sulfoaldenafil, an ingredient whose adverse effects on humans were never tested. It is estimated that more than 50,000 new dietary supplements have been available on the market since the DSHEA became law [2]. DSHEA implementation addressed dietary supplement safety concerns by requiring manufacturers to provide adequate proof to the FDA that the new ingredients introduced are safe [4, 5]. The regulation was poorly enforced, as the FDA received appropriate notifications for only 170 new ingredients [2] potentially making consumers more vulnerable to these potential supplement adverse events.

Literature review has been used to study the adverse events of dietary supplements. Pittler, Schmidt, and Ernst conducted a systematic review on the adverse events of herbal food supplements for body weight reduction including 33 case reports and 48 clinical trials dating from 1995 to 2004 [7]. Seven ingredients were covered in this review and their adverse events were reported in detail. However, literature-based studies suffer from several drawbacks. Firstly, literature-based reviews can only cover a few dietary supplements. In the example above, the authors could only focus on seven ingredients intended for body weight reduction, which is only a tiny portion of the dietary supplement ingredients currently available on the market. Secondly, the qualities of the literature are inconsistent. The literature that reported the adverse events did not explain the related mechanism or establish any cause and effect relationship. Finally, incompleteness of the literature is an intrinsic limitation of literature-based studies [7, 8].

Alternatively, dietary supplement product labels contain vital information on the adverse events of dietary supplements. The FDA has stringent rules on dietary supplement labeling information including statement of identity, active ingredients, manufacturer information, structure or function claims, and health claims to make sure the information is truthful and not misleading [4,5,9]. Consequently, the supplement labels provide an opportunity for more efficient methods of supplement adverse events related research. However, unlike drug labeling, adverse events related to the supplements typically exist in the precaution section of labels in free text format. Thus, text mining techniques are required to extract adverse events information from supplement labels.

Topic modeling is a set of text mining methods that automatically discover the underlying themes of a collection of documents without prior document annotation or labeling. Implementations of topic modeling includes latent semantic indexing (LSI), probabilistic latent semantic analysis (PLSA) [10], and latent Dirichlet Allocation (LDA). LDA is one of the more common topic model techniques in the current literature. A topic is defined as a “distribution over a fixed vocabulary” [11]. LDA is a statistical model, which assumes that all the

documents in the collection can be described by a group of topics, but each document is a different distribution of these topics. As a generative graphic model, LDA can discover the underlying topic distribution for a large document collection.

Topic modeling has been applied to social media and drug product labeling to discover new knowledge. Sullivan *et al.* have recently applied LDA to Amazon.com nutritional supplement reviews to find potentially unsafe dietary supplements [12]. Bisgin *et al.* have used LDA on FDA drug labels to group drugs with similar safety concerns and therapeutic uses [13], and this information was later used to discover drug repositioning opportunities [14]. However, the methodology has to the best of our knowledge not been applied to dietary supplement labels. Product labels provide useful information to group dietary supplements based on the listed adverse events. Thus, in this study, we tried to demonstrate that the application of the LDA technique on the precautions statements of dietary supplement labels can yield useful groupings of adverse events without any prior knowledge to analyze the adverse events.

Dietary Supplement Label Database (DSLID)

The DSLID is created and managed by Office of Dietary Supplement (ODS) and U.S. National Library of Medicine (NLM) in the National Institutes of Health (NIH) [15]. The database includes complete label contents among 55,456 dietary supplement products currently marketed, off the market, and consumed by NHANES participants. For each supplement product, there are four distinct label sections: namely, product information (including product name, statement of identity, serving information, and target groups); dietary supplement facts (including active ingredients); label statements (including formulation, precaution, and suggested use); and contact information (including manufacturer information).

DSLID provides a web application programming interface (API) for an efficient extraction of the labeling information in JavaScript Object Notation (JSON) format. It also provides a comma-separated value (CSV) file including database identifiers (DSLID ID) for all the products in the database.

Side Effect Resource (SIDER)

The SIDER database is a free resource that contains 5,868 adverse drugs reactions (ADRs) of 1430 drugs. It uses the Medical Dictionary for Regulatory Activities (MedDRA) terms to describe the drug side effects. The list of side effect MedDRA terms was extracted from the drug labels and downloaded in CSV using SIDER version 4.1 [16].

Although the SIDER database is not designed for dietary supplements, the adverse events encountered with the dietary supplements are similar to those with drugs. Therefore, the SIDER MedDRA terms list was used to recognize adverse event keywords in the DSLID labels.

Methods

Figure 1 illustrates the overview of the methods. We extracted the supplement label information from the DSLID, preprocessed the label statements to create a collection of supplement documents represented by a list of SIDER MedDRA terms, and then grouped supplements using LDA. The adverse events topics generated by LDA were further evaluated by human annotation.

DSLID Label Preprocessing

The labels of 55,456 dietary supplements available in DSLID were extracted via the API in JSON format. Two criteria were

applied to generate a subset of dietary supplements products for topic modeling analysis.

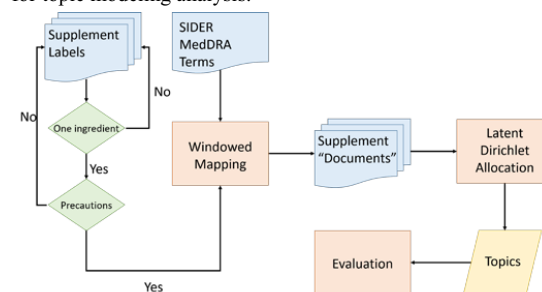


Figure 1—Method overview flow chart

1. Products that only have one active ingredient entry in the dietary supplement facts section were selected for further preprocessing to prevent interactions between ingredients from confounding our analysis.
2. Products that have one or more statements under the “Precautions” subsection of the label statements section.

The “Precautions” statements of each supplement in the subset were split into a list of single sentences with the punctuation marks removed and all words lowercased. We mapped terms (allowing a window size of 5 for multiple-token terms) to the MedDRA terms listed in SIDER for each supplement product. A windowed mapping was used since a multiple-token phrase may be expressed noncontiguously or in different order. For example, the following statement, “Beta-alanine may cause a harmless, temporary tingling or flushing sensation”, should match the word “tingling sensation” in the MedDRA terms list although the two words do not appear next to each other. A windowed mapping will yield the correct result, while an exact match will not.

It is noteworthy that although the supplements in the dataset all have only one active ingredient entry, some of the ingredients are actually blends. For example, “chocamine” was registered as a single ingredient, but in fact it consists of a mixture including theobromine, caffeine, theophylline, phenylethylamine, tyramine, phenylalanine, tryptophan, and tyrosine. This type of supplements was excluded in our analysis due to the complexity of the components in blends.

The windowed mapping generates a “document”, which is a list of the MedDRA terms appeared in the “Precautions” statement for each supplement. In other words, each supplement is represented by a list of MedDRA terms. The collection of the documents in the subset was then analyzed by LDA model.

Topic Modeling Analysis

The LDA model was implemented using the Python *gensim* package [17]. Fifty topics were generated by the training process using the dataset of preprocessed labels. To simplify the analysis, we assigned each supplement only one topic. Each topic was represented by the topic keywords, i.e. MedDRA terms, and their corresponding probabilities. As discussed in the introduction, a topic is a distribution over all the MedDRA terms which appeared in the label statements. Therefore, to better define a topic in terms of adverse events, the topic keywords with probabilities lower than 0.05 were not considered to define a topic. Due to the probabilistic nature of the LDA model, the training was repeated 10 times to see if there are significant differences among the generated topics.

The conditional probability of each topic given a particular supplement was calculated, and the supplement was assigned to the topic that has the maximal conditional probability. The distribution of the supplements over the topics was then derived by counting the number of supplements under each topic.

Topic Evaluation

Since the topics generated by LDA are not based on any prior annotation or labeling of the statements, it is necessary to evaluate the topics by checking if the supplements assigned to a certain topic are actually related to the adverse events the topic represents.

Two criteria were used to select the topics to be evaluated:

1. The chosen topics all have more than one keyword with probabilities greater than 0.05. This study is focused on grouping of dietary supplements by their adverse events. If a topic is defined by more than one adverse event, then it is necessary to examine if these adverse events are related in terms of supplement function.
2. The number of supplements under the chosen topics should be within the range of 10-50. If there are too few products under a certain topic, there would be insufficient data points to generate any significant grouping patterns; on the other hand, if there are too many products under a certain topic, the grouping may not be meaningful, as a group may contain a mixture of completely different supplements.

Eight topics out of 50 were chosen accordingly for evaluation based on the group of adverse events indicated by the topic keywords. The largest topic contained 46 products, while the smallest topic contained 13 products.

Two human annotators manually reviewed the active ingredients and the health claims in the statements on each supplement labels under the eight topics. Annotators made a consensus on their annotations when there was a disagreement on a product assigned for a given topic. To quantitatively evaluate the performance of topic modeling, we used the metric called topic relatedness, defined as the percentage of products (or ingredients) related to the topic keywords out of the total number of products (or ingredients) assigned to the same topic.

Results

DSLID Label Preprocessing

There are 6,123 side effect terms that were extracted from the drug labels in the MedDRA term list. Among 55,456 dietary supplements available in DSLID, 27.9% (15,452) of the supplement products have only one active ingredient entry. Only 3.6% (2,014) of these single ingredient supplements whose "Precautions" label statements contained adverse events were matched to a total of 3.9% (239) terms within the MedDRA term list.

Topic Modeling Analysis

The distribution of the 2,014 supplements over the 50 topics generated by the LDA was shown in Figure 2. The topic with the most supplements has 234 products, while the topic with the least supplements has only 1 product. The median of the supplement counts over the 50 topics is 21, and the first quartile and the third quartile is 8 and 39, respectively.

The top three topics in supplement counts grouped 663 (32.9%) supplements that have precautions related to pregnan-

cy and breast feeding, and all of these three topics have only a single topic keyword to represent their adverse events. However, these topics did not satisfy the criteria for further evaluation. Therefore, another eight topics including 210 products were selected for evaluation as shown in Table 1. The number of one-ingredient products and the unique ingredients of each topic was shown in Table 2.

Topic Evaluation

The topic relatedness of each topic on the product and the ingredient level was shown in Figure 3. Topic 24, 29, and 47 have reached a topic relatedness of 100% at both product level and ingredient level. This suggests that all the supplements grouped under these topics have mentioned the adverse events given in the topic keywords in their label statements.

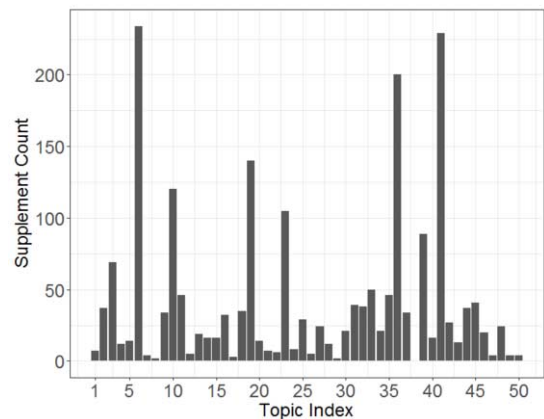


Figure 2– The distribution of the supplement counts for the 50 topics

Table 1- Topic keywords and the representative ingredients of eight topics chosen for manual evaluation. The topic keywords were listed in decreasing order by their probabilities generated by LDA. The bolded topics are coherent ones where the topic keywords are closely related.

Topic	Topic Keywords	Representative Ingredients
1	Sarcoidosis, hyperparathyroidism, lymphoma, bipolar disorder	Vitamin D3
11	Anxiety, depression, anxiety depression, anxiety disorder	S-Adenosyl-L-Methionine (SAME), steroids
12	Obstruction, inflammation, intestinal obstruction, pain, abdominal pain	Cascara sagrada, Celandine
24	Hyperthyroidism, palpitations, duodenal ulcer, anxiety, blood pressure abnormal	L-Tyrosine, Iodine
29	Gastritis, loose stools, peptic ulcer, heartburn, ulcer	Oregano oil, Cayenne pepper
33	Photosensitivity, tingling skin, tingling sensation, bladder dysfunction, serotonin syndrome	Beta-Alanine, St. John's Wort
42	Epilepsy, blood disorder,	St. John's Wort,

47	cardiovascular disorder, hypotension, tuberculosis	Graviola
	Dizziness, headache, nervousness, glaucoma, depression	Caffeine, L- Glutamine

The topic relatedness of the remaining five topics indicates that there exist supplements that do not contain statements relevant to the adverse events mentioned in the corresponding topic keywords. The discrepancies between the product and the ingredient level relatedness are also different among the topics. For Topic 1, 33, and 42, the product level topic relatedness is higher than the ingredient level one. The opposite is true for Topic 11 and 12. Topic 12 had the lowest product level topic relatedness at 58.8%, and Topic 33 had the lowest ingredient level topic relatedness at 57.1%.

Table 2- Comparison of the number of one-ingredient products and the number of unique ingredients in the topics under evaluation.

Topic	One-ingredient products	Unique Ingredients
1	37	8
11	44	34
12	17	6
24	26	7
29	18	6
33	20	7
42	12	10
47	19	12

The inconsistent differences for topic relatedness between product level and ingredient level relatedness among different groups are due to the fact that some products in the same group share the same ingredients. If there exist several products containing the same ingredient that are related to the adverse events in the topic, a lower topic relatedness at ingredient level is expected. For example, there are 22 products that contain vitamin D3 under Topic 1 and vitamin D3 was deemed a related ingredient that may cause these adverse events. Therefore, a lower ingredient level relatedness for this topic was obtained.

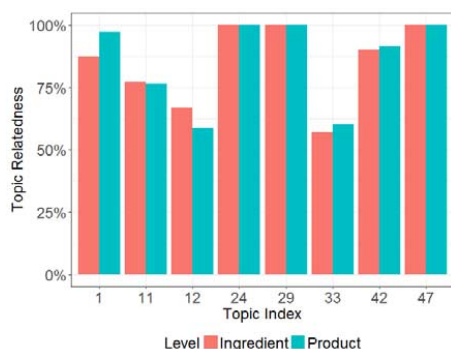


Figure 3- Topic relatedness at both the product and the ingredient level of the eight topics listed in Table 1. The x-axis is labeled with the topic index.

Conversely, if there exist several products containing the same ingredient that are irrelevant to the adverse events in the topic, a higher ingredient level topic relatedness will be observed. The opposite example could be found in the topic related to obstruction, inflammation, intestinal obstruction, pain, and

abdominal pain. Four products containing celandine were grouped under Topic 12, but celandine is an ingredient that was judged as not related to the adverse events mentioned in the topic. As a result, a higher ingredient level topic relatedness was observed.

Discussion

The adverse effects of dietary supplements were brought under more careful scrutiny due to the rapid growth of the supplement market and consumer base. The special regulatory framework for dietary supplements has led to both challenges and opportunities for research in supplement adverse events. The challenge lies in the fact that the soaring number of new supplement products introduced to the market demands more efficient ways to study the adverse events of a wide variety of dietary ingredients, while the opportunity arises in the supplement labels where relevant information of adverse events was provided as strictly required by the FDA. Topic modeling, specifically LDA, is a powerful method to find hidden grouping patterns within a series of documents without any prior knowledge or annotation. Therefore, the objective of this study is to utilize LDA as an effective method to generate useful groupings of dietary supplements based on the adverse events found in their supplement labels.

Since each topic generated by LDA is represented by a list of adverse events, we evaluated a subset of these topics by human annotation whether the supplements under each topic were actually related to the adverse events indicated by the topic keywords. If a topic is represented by multiple adverse events, the relationship among them was further examined to determine if there are multiple sub-topics.

Our evaluation results have shown that LDA was able to find similar supplement products and categorize them into the same topic based only on the adverse events mentioned in product labels. These supplements may contain the same active dietary ingredient or different ingredients with similar functions. For instance, in Topic 11, four different products with the same ingredients SAME and six products containing dietary ingredients that are all different derivatives of androstosterone were grouped together due to their potential risk of causing anxiety and depression.

We also found that LDA could associate clinically related adverse events with each other within the same topic. Among the eight topics listed in Table 1, five were highlighted in bold because the adverse events representing each topic were related and forming a single coherent topic. For example, in Topic 1, hyperparathyroidism, lymphoma, and sarcoidosis are able to cause hypercalcemia [18]. Bipolar disorder patients who are taking lithium medication are at risk of lithium-induced hypercalcemia [19]. This suggests that the dietary supplements reviewed under this topic were all linked to calcium metabolism. Although hypercalcemia did not appear in the dietary supplement labels, it is the hidden commonalities among the four adverse events that represent the topic. The most frequent ingredients that appeared under this topic is vitamin D3, which is an important factor in the calcium metabolism [20]. This suggests that consumers with any one of the pre-existing conditions mentioned in the topic may expose themselves to adverse events by taking vitamin D3. Thus, the above examples demonstrated that LDA had successfully grouped dietary supplements based only on their possible adverse effects listed in the product labels into a coherent topic. The topics can give insight into further literature-based studies on the adverse effects of a particular set of dietary supplements.

The study has its limitations. Firstly, the MedDRA terms included in the SIDER does not cover all possible synonyms

that may appear in the label statements. For example, loss of hair or hair loss, which are the synonyms for alopecia, are not included in the MedDRA terms list, but were present in supplement labels. The windowed mapping method requires every word in the MedDRA terms to be present in the statements, and therefore was not able to capture all the variations of the same adverse effects present in the labels.

Since LDA is an unsupervised machine learning method based only on the word frequencies in a document, the topic model cannot differentiate supplement indications from adverse events of supplement as it does not take context into account. For example, in Topic 12, *Cascara sagrada* was used as an anthranoid laxative to treat constipation [21], and celandine was indicated for spastic discomfort of the gastrointestinal tract [22]. The topic keywords, obstruction, intestinal obstruction, abdominal pain were not adverse events, but indications of these herbal supplements.

Our LDA analysis only considered the supplements with only one active ingredient entry. According to DSLD, more than 70% of the supplements contain at least two active ingredients. Therefore, it is of our interest to analyze this portion of the dietary supplements while addressing the entity recognition issue (the indications, the adverse events, and the interactions between ingredients) in the future. We may consider using existing knowledge base content to find the known interactions of ingredients in products.

Conclusion

In summary, we extracted the precautions statements from the labeling information of 2,014 dietary supplements that have only one active ingredient in DSLD. The MedDRA terms list was used to convert the statements into documents represented by a list of adverse events. The collection of supplement documents were analyzed by LDA topic models and eight out of the 50 resulting topics were evaluated by two human annotators. The product level topic relatedness ranged from 58.8% to 100%, while the ingredient level topic relatedness ranged from 57.1% to 100%. Five topics have shown considerable coherence among the topic keywords, and the representative ingredients have been demonstrated to be closely related to the topics. These results indicated that LDA could effectively group the dietary supplements by their adverse effects, and the grouping information could provide insight for further literature-based studies.

Acknowledgements

Research reported in this publication was supported by the National Center for Complementary & Integrative Health Award (R01AT009457) (Zhang) and the University of Minnesota Grant-In-Aid award (Zhang).

References

- [1] E.D. Kantor, C. D. Rehm, M. Du, E. White, and E. L. Giovannucci, Trends in Dietary Supplement Use among US Adults from 1999-2012, *JAMA* **316** (2016), 1464-1474.
- [2] P. A. Cohen, Assessing Supplement Safety – The FDA’s Controversial Proposal, *N Engl J Med* **366** (2012), 389-391.
- [3] A. I. Geller, N. D. Shehab, N. J. Weidle, M. C. Lovegrove, et al., Emergency Department Visits for Adverse Events Related to Dietary Supplements, *N Engl J Med* **373** (2015), 1531-1540.
- [4] M. J. Mason, Drugs or Dietary Supplements: FDA’s Enforcement of DSHEA, *J Public Policy Mark* **17** (1998), 296-302.
- [5] V. H. Frankos, D. A. Street, and R. K. O’Neill, FDA Regulation of Dietary Supplements and Requirements Regarding Adverse Event Reporting, *Clin Pharmacol Ther* **87** (2009), 239-244.
- [6] B. E. Denham, Dietary Supplements – Regulatory Issues and Implications for Public Health, *JAMA* **306** (2011), 428-429.
- [7] M. H. Pittler, K. Schmidt, E. Ernst, Adverse events of herbal food supplements for body weight reduction: systematic review, *Obes Rev* **6** (2005), 93-111.
- [8] A. A. Izzo, E. Ernst, Interactions between herbal medicines and prescribed drugs: an updated systematic review, *Drugs* **69** (2009), 1777-1798.
- [9] FDA regulations on supplement labeling. Available from: <http://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/DietarySupplements/default.htm>
- [10] L. Liu, L. Tang, W. Dong, S. Yao, and W. Zhou, An overview of topic modeling and its current applications in bioinformatics, *SpringerPlus* **5** (2016), 1608.
- [11] D. M. Blei, Probabilistic Topic Models, *Commun ACM* **55** (2012), 77-84.
- [12] R. Sullivan, A. Sarker, K. O’Connor, A. Goodin, M. Karlsrud, and G. Gonzalez, Finding potentially unsafe nutritional supplements from user reviews with topic modeling, *Pac Symp Biocomput* **21** (2016), 528-39.
- [13] H. Bisgin, Z. Liu, H. Fang, X. Xu, and W. Tong, Mining FDA drug labels using an unsupervised learning technique – topic modeling, *BMC Bioinformatics* **12** (2011), S11.
- [14] H. Bisgin, Z. Liu, R. Kelly, H. Fang, X. Xu, W. Tong, Investigating drug repositioning opportunities in FDA drug labels through topic modeling, *BMC Bioinformatics* **13** (2012), S6.
- [15] Dietary Supplement Label Database. Available from: <https://dslid.nlm.nih.gov/dslid>
- [16] M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork, The SIDER database of drugs and side effects, *Nucleic Acids Res* **44** (2016), D1075-D1079.
- [17] R. Rehurek, and P. Sojka, Software Framework for Topic Modelling with Large Corpora, *Proc. LREC 2010 Workshop on New Challenges for NLP Frameworks*, Valletta, Malta, 2010.
- [18] P. K. Pellitteri, Evaluation of hypercalcemia in relation to hyperparathyroidism, *Otolaryngol Clin North Am* **43** (2010), 389-397.
- [19] M. E. Wolf, M. Moffat, J. Mosnaim, and S. Dempsey, Lithium therapy, hypercalcemia and hyperparathyroidism, *Am J Ther* **4** (1997), 323-325.
- [20] H. F. DeLuca, The metabolism and functions of vitamin D, *Adv Exp Med Biol* **196** (1986), 361-75.
- [21] E. S. Nakasone, and J. Tokeshi, A Serendipitous Find: A Case of Cholangiocarcinoma Identified Incidentally After Acute Liver Injury Due to *Cascara sagrada* Ingestion, *Hawaii J Med Public Health* **74** (2015), 200-202.
- [22] R. Teschke, C. Frenzel, X. Glass, J. Schulze, and A. Eickhoff, Greater Celandine hepatotoxicity: a clinical review, *Ann Hepatol* **11** (2012), 838-48.

Address for correspondence

Rui Zhang, PhD, University of Minnesota, Minneapolis, MN, USA
zhan1386@umn.edu

An Early Infectious Disease Outbreak Detection Mechanism Based on Self-Recorded Data from People with Diabetes

Ashenafi Zebene Woldaregay^a, Eirik Årsand^b,
Taxiarchis Botsis^a, Gunnar Hartvigsen^{a, b}

^a Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway,

^b Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

Abstract

People with diabetes experience elevated blood glucose (BG) levels at the time of an infection. We propose to utilize patient-gathered information in an Electronic Disease Surveillance Monitoring Network (EDMON), which may support the identification of a cluster of infected people with elevated BG levels on a spatiotemporal basis. The system incorporates data gathered from diabetes apps, continuous glucose monitoring (CGM) devices, and other appropriate physiological indicators from people with type 1 diabetes. This paper presents a novel approach towards modeling of the individual's BG dynamics, a mechanism to track and detect deviations of elevated BG readings. The models were developed and validated using self-recorded data in the non-infection status using Dexcom CGM devices, from two type 1 diabetes individuals over a 1-month period. The models were also tested using simulated datasets, which resemble the individual's BG evolution during infections. The models accurately simulated the individual's normal BG fluctuations and further detected statistically significant BG elevations.

Keywords:

Blood Glucose; Disease Outbreaks; Diabetes Mellitus, Type 1

Introduction

Diabetes mellitus is a chronic disease that causes blood glucose (BG) metabolic disorder [1], either due to failure of the pancreas beta cells to produce insulin (type 1) or failure of the body to react to insulin in the proper way (type 2) [2]. Many people with diabetes strive to control their BG levels as close to the normal range as possible to avoid medical complications. In this regard, many diabetes self-management applications and devices have been developed to support this patient group, of which almost all of them take the ubiquitous nature of mobile devices as an advantage to base their development [3-6]. Moreover, diabetes self-management applications and devices have shown feasible to integrate with the patient's Electronic Health Record system [7-9]. The advent of information technology and availability of biosensors and point of care technologies have also paved the way for quantified self and an easy near patient testing [10; 11]. These advancements have further enhanced the opportunity of using the individual's diabetes data and other physiological indicators for secondary purposes.

People with diabetes experience elevated BG levels when in infectious state [12; 13]. A positive correlation between high BG levels and infections has been demonstrated in the cases of Influenza, Cholera, Plague, Ebola, Anthrax and SARS [14; 15]. Use of BG levels for an early outbreak detection has been

suggested in works of literature [13; 14; 16-19]. For example, Granberg et al. [19] introduced an automatic infection detection system based on the individual's BG levels. Årsand et al. [14] described the system architecture, model and requirements of disease surveillance based on patient observable parameter, i.e. blood glucose. Botsis et al. [20] assessed the development of electronic disease surveillance systems for detecting infections at the early stages, i.e. during the incubation period. Aside from these potential methods, no practical way of facilitating this has been identified.

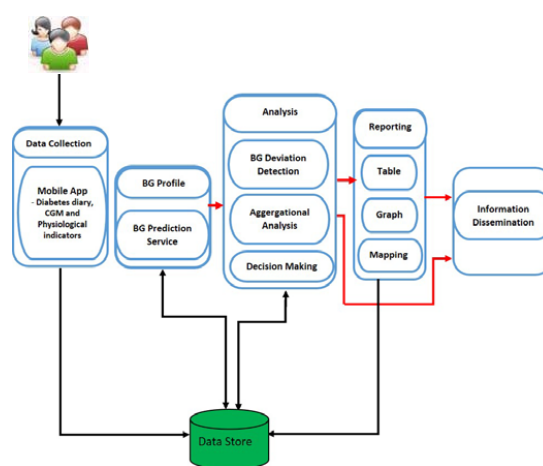


Figure 1 - Proposed EDMON System Architecture.

We propose the development of an Electronic Disease Surveillance Monitoring Network (EDMON), which may support the detection of infections before the onset of the first symptoms. The system incorporates data gathered from diabetes apps, CGM devices, and other appropriate physiological indicators from people with type 1 diabetes. It consists of five different modules, as shown in Figure 1, with different functionalities; data collection, BG profile, analysis, reporting, and information dissemination modules. The task of the data collection module is to gather the individual's diabetes data and other physiological indicators from diabetes apps, wearables, and sensors, whereas BG profile module will make a prediction of individuals' upcoming BG levels based on past status. The analysis module will analyze the individual's physiological indicators and compare the predicted and actual BG levels for

any statistically significant deviations. It also aggregates and determines the presence of any aberrant pattern within a cluster of people based on a spatiotemporal basis. If outbreak is detected, the reporting module will prepare the information in a way suitable for the end users, whereas the information dissemination module will handle the delivery of the information for the end users. Generally, the objective of the EDMON project is to design and develop an electronic disease monitoring network based on inputs from people with diabetes, which can track real-time BG levels of each individual independently and detect a cluster of infected people with statistically significant BG deviations on a spatiotemporal basis. This paper presents the first step of the EDMON development with modeling of the individual's BG dynamics. We consider the development of the personalized model as the core part of the EDMON system and describe the details of our approach in the next sections.

Materials and Methods

Materials

The models were developed using one-month BG data from two type 1 diabetes people, as shown in Figure 2, sampled using Dexcom CGM devices. The actual self-recorded data were used to develop and validate the model fit with individuals' BG dynamics in the non-infection state. We also generated simulated datasets that resemble the individual's BG fluctuations during infections, as shown in Figure 3. The simulated datasets included unexpected elevated BG levels for a certain period of time with some random and steady increments per hours ($\Delta BG/\text{hour}$). The unexpected BG variations were defined as 1) any unstructured variations outside of the individual's modeled structured variations and, 2) inherent stochastic phenomena of the BG dynamics that cannot be quantified with the developed models using the key diabetes parameters (e.g. BG, insulin intake, physical activities, and dietary habits). The simulated datasets were used to test the model's performance in capturing the assumed infection-related, unexpected, elevated BG readings. The algorithm was developed using Matlab version R2015b.

Methods

Our models combined a novel approach for BG monitoring and outlier detection, which was based on a set of autoregressive models and predicts the individuals' expected BG values on an interval basis. The actual BG value was compared with the predicted intervals, which was generated using autoregressive (AR) model [21], Eq. (1), and autoregressive moving average (ARMA) model [21; 22], Eq. (2).

$$y(n) = \sum_{i=1}^q \mathcal{A}(i)y(n-i) + e(n) \quad (1)$$

Where, $\mathcal{A}(q)$ are autoregressive coefficients, and $y(n)$ is n^{th} BG value.

$$\sum_{i=0}^q \mathcal{A}(i)y(n-i) = \sum_{i=0}^q \mathcal{C}(i)e(n-q) \quad (2)$$

Where, $\mathcal{A}(q)$ are autoregressive coefficients, $\mathcal{C}(q)$ are moving average coefficients, and $y(n)$ is the n^{th} BG value, and $e(n)$ are Gaussian noise with zero mean and constant variance σ_e^2 .

The prediction intervals [23-25] were computed based on the empirical distribution of errors between the predicted and ac-

tual BG values for the prediction horizon under consideration, using Eq. (3).

$$\tilde{y}(n) = y(n) \pm z_{\alpha/2} \sqrt{\text{Var}[e_k(n)]} \quad (3)$$

Where, $\tilde{y}(n)$ is the predicted BG intervals, $y(n)$ is the model's point BG prediction, $z_{\alpha/2}$ is the assumed errors distribution, α is level of significance, and $\text{Var}[e_k(n)]$ is variance of the errors for a specific window size, w . The prediction interval was computed and compared to various values of window size and level of significance. The optimal prediction interval was reported with a window size (w) and level of significance (α).

MATLAB system identification toolbox along with partial autocorrelation function (PACF) were used to identify the optimal model order. The empirical distribution of errors between the actual and predicted values were assumed to follow a normal distribution.

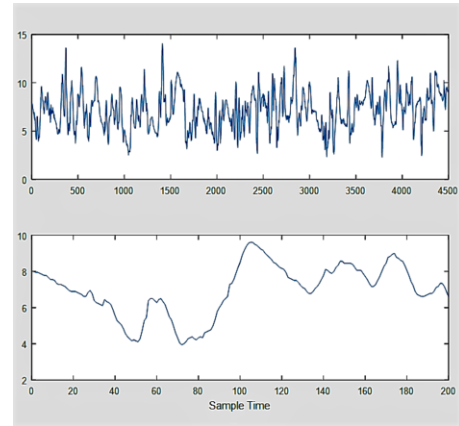


Figure 2 - Plot of the CGM data. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

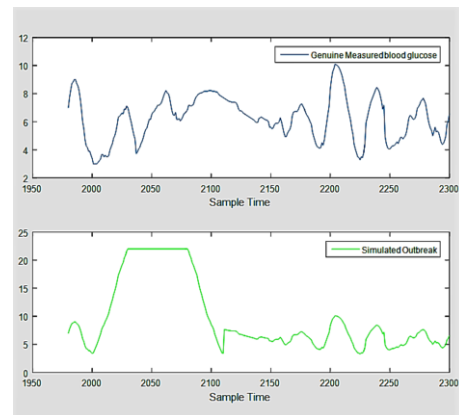


Figure 3 - Simulated BG values in response to infections (Top -non-infectious BG levels, Bottom -simulated infection related BG levels). [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Results

Autoregressive (AR) Model

For both the first and the second subject, an autoregressive (AR) model of order five ($p = 5$) was found to be optimal for the point BG prediction and fitted best with a root mean square error (RMSE) of 0.2159 and 0.3068 respectively. For the first subject, the predicted interval was found to be optimal with a window size of $w = 100$ and a statistically significant level of $\alpha = 0.01$, see Figure 4. The predicted interval for the second subject was also found to be optimal with a window size of $w = 200$ and a statistically significant level of $\alpha = 0.01$, see Figure 5.

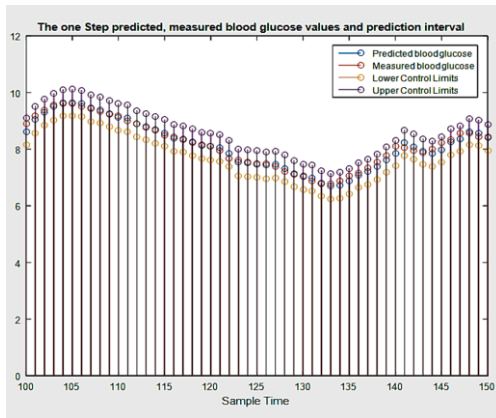


Figure 4 - Subject one- autoregressive (AR) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

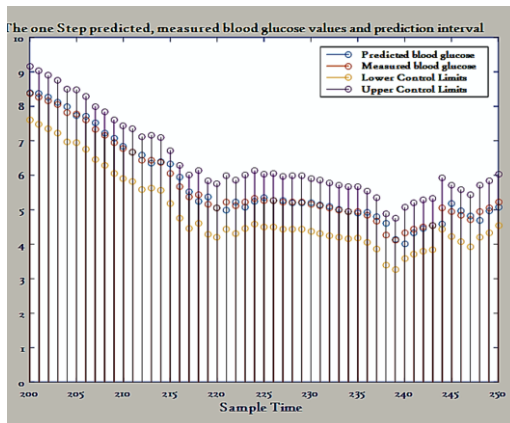


Figure 5 - Subject two - autoregressive (AR) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Autoregressive Moving Average (ARMA) Model

For both subjects, the optimal autoregressive moving average (ARMA) model order was found to be an autoregressive order of 6, and a moving average order of 2. The point BG prediction for the first subject resulted in a root mean square error (RMSE) of 0.2114. The predicted interval was found to be

effective with a window size of $w = 100$ and a statistically significant level of $\alpha = 0.01$, see Figure 6. The point BG prediction for the second subject also resulted in a RMSE of 0.2915. The predicted interval was found to be effective with a reasonable window size of $w = 200$ and a statistically significant level of $\alpha = 0.01$, see Figure 7.

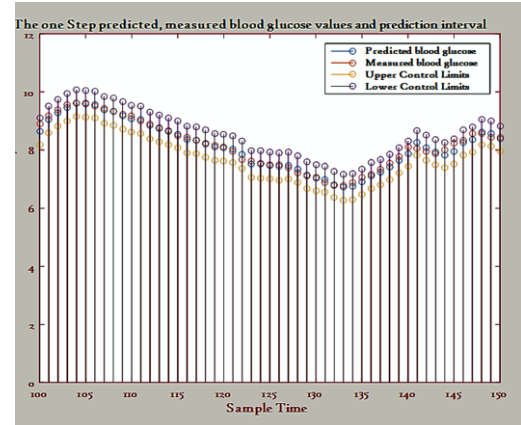


Figure 6 - Subject one - autoregressive moving average (ARMA) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

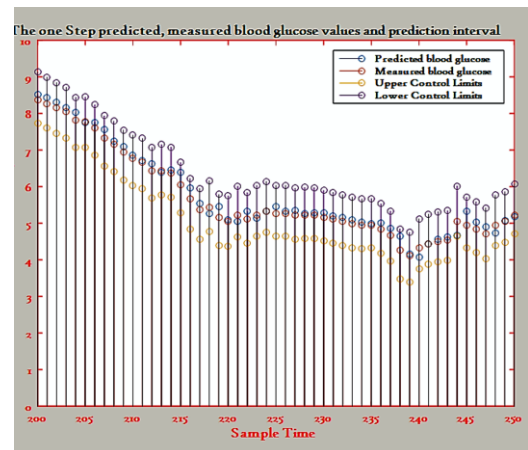


Figure 7 - Subject two- autoregressive moving average (ARMA) model. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

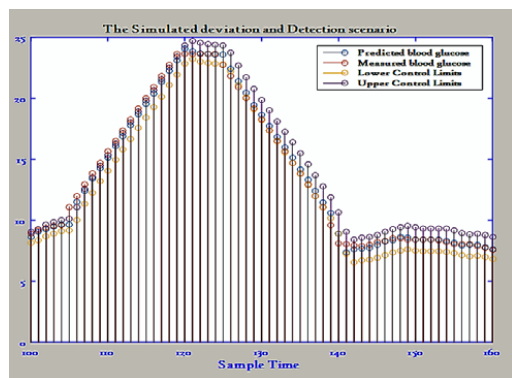


Figure 8 - Detection capabilities of the developed algorithm on testing data. [x and y-axes represent CGM's sampling time and BG level in mmol/l, respectively].

Deviation Detection/Surveillance

The algorithm was tested with simulated datasets, for its capability of detecting unexpected BG variations that may occur during infections, as shown in Figure 3. The developed algorithm successfully detected statistically significant elevated BG readings, as shown in Figure 8. As clearly shown in the Figure, the algorithm was highly sensitive to the slope, and clearly captured the rise and fall of the individual's BG readings.

Discussion

The recent advancement, syndromic surveillance uses health-related data that precede diagnosis and laboratory verification to produce signals with sufficient probability of outbreaks to warrant further actions [26-30]. The development of strategies for early detection of outbreaks is worthy, given the limitations of the existing disease surveillance systems. We presented a modelling approach for the early detection of infections in people with diabetes. This set of models were developed as part of the EDMON system that will rely on real-time data collection from people with diabetes. Our approach was capable of detecting statistically significant and unexplained BG elevations of various size and duration.

One of the limitations of our study is the sample size. We also have used only BG as the input variable. Furthermore, the assumption of a normal error distribution could be a limitation, which needs further exploration. To alleviate these limitations, we plan to explore other more robust approaches and involve real infection related BG data, more input variables (insulin intake, physical activity, and diet) along with a larger sample size.

Conclusion

EDMON is an early outbreak detection system that relies on self-gathered health-related data from people with type 1 diabetes. In this paper, we presented a novel approach that can track BG levels and detect statistically significant BG elevations. The testing and validation of this approach on a large-scale basis could support the development of an outbreak detection system based on real-time data collection from people with diabetes. We believe such efforts may lay the foundations for the next generation disease surveillance systems and provoke further thoughts in this valuable field.

References

- [1] W.H. Organization, Diabetes Fact Sheet, in, 2016.
- [2] R.L. Richesson, S.A. Rusincovitch, D. Wixted, B.C. Batch, M.N. Feinglos, M.L. Miranda, W.E. Hammond, R.M. Califf, and S.E. Spratt, A comparison of phenotype definitions for diabetes mellitus, *J Am Med Inform Assoc* **20** (2013), e319-326.
- [3] E. Årsand, N. Tataru, G. Østengen, and G. Hartvigsen, Mobile Phone-Based Self-Management Tools for Type 2 Diabetes: The Few Touch Application, *Journal of Diabetes Science and Technology* **4** (2010), 328-336.
- [4] D.-Z. Issom, A.Z. Woldaregay, T. Chomutare, M. Bradway, E. Årsand, and G. Hartvigsen, Mobile applications for people with diabetes published between 2010 and 2015, *Diabetes Management* **5** (2015), 539-550.
- [5] C.C. Quinn, S.S. Clough, J.M. Minor, D. Lender, M.C. Okafor, and A. Gruber-Baldini, WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction, *Diabetes Technol Ther* **10** (2008), 160-168.
- [6] K. Waki, H. Fujita, Y. Uchimura, K. Omae, E. Aramaki, S. Kato, H. Lee, H. Kobayashi, T. Kadowaki, and K. Ohe, DialBetics: A Novel Smartphone-based Self-management Support System for Type 2 Diabetes Patients, *J Diabetes Sci Technol* **8** (2014), 209-215.
- [7] P.-Y. Benhamou, Improving diabetes management with electronic health records and patients' health records, *Diabetes & Metabolism* **37** (2011), S53-S56.
- [8] T.C. Veinot, K. Zheng, J.C. Lowery, M. Souden, and R. Keith, Using Electronic Health Record Systems in Diabetes Care: Emerging Practices, *IHI ... : proceedings of the ... ACM SIGHT International Health Informatics Symposium. ACM SIGHT International Health Informatics Symposium 2010* (2010), 240-249.
- [9] O.A. Walseth, E. Årsand, T. Sund, and E. Skipenes, Wireless Transfer of Sensor Data into Electronic Health Records, in: *Connecting Medical Informatics and Bio-Informatics*, IOS Press, 2005.
- [10] T. Botsis, S. Walderhaug, A. Dias, K. van Vuurden, J.G. Bellika, and G. Hartvigsen, Point-of-care devices for healthy consumers - a feasibility study, *J Telemed Telecare* **15** (2009), 419-420.
- [11] B.C. Delaney, C.J. Hyde, R.J. McManus, S. Wilson, D.A. Fitzmaurice, S. Jowett, R. Tobias, G.H. Thorpe, and F.D. Hobbs, Systematic review of near patient test evaluations in primary care, *BMJ* **319** (1999), 824-827.
- [12] S. Knapp, Diabetes and infection: is there a link?—A mini-review, *Gerontology* **59** (2013), 99-104.
- [13] J.N. Lauritzen, E. Årsand, K. van Vuurden, J.G. Bellika, O.K. Hejlesen, and G. Hartvigsen, Towards a mobile solution for predicting illness in Type 1 Diabetes Mellitus: Development of a prediction model for detecting risk of illness in Type 1 Diabetes prior to symptom onset, (2011), 1-5.
- [14] E. Årsand, O.A. Walseth, N. Andersson, R. Fernando, O. Granberg, J.G. Bellika, and G. Hartvigsen, Using Blood Glucose Data as an Indicator for Epidemic Disease Outbreaks, *Connecting Medical Informatics and Bio-Informatics* **116** (2005), 217-222.
- [15] T. Botsis, O. Hejlesen, J.G. Bellika, and G. Hartvigsen, Electronic disease surveillance for sensitive population groups - the diabetics case study, *Stud Health Technol Inform* **136** (2008), 365-370.
- [16] T. Botsis, O. Hejlesen, J.G. Bellika, and G. Hartvigsen, Blood glucose levels as an indicator for the early detection of infections in type-1 diabetics, in: *Advances in Disease Surveillance*, 2007, p. 147.
- [17] T. Botsis and G. Hartvigsen, Exploring new directions in disease surveillance for people with diabetes: lessons learned and future plans, *Stud Health Technol Inform* **160** (2010), 466-470.
- [18] A.Z. Woldaregay, K. van Vuurden, E. Årsand, T. Botsis, and G. Hartvigsen, Electronic Disease Surveillance System Based on Input from People with Diabetes: An Early Outbreak Detection Mechanism, in: *Proceedings from The 14th Scandinavian Conference on Health Informatics 2016, Gothenburg, Sweden, April 6-7 2016*, Linköping University Electronic Press, 2016, pp. 23-27.
- [19] O. Granberg, J.G. Bellika, E. Årsand, and G. Hartvigsen, Automatic infection detection system, *Stud Health Technol Inform* **129** (2007), 566-570.
- [20] T. Botsis, J.G. Bellika, and G. Hartvigsen, New Directions in Electronic Disease Surveillance: Detection of Infectious Diseases during the Incubation Period, *International Conference on Ehealth, Telemedicine, and Social Medicine: Etelemed 2009, Proceedings* (2009), 176-183.
- [21] M. Eren-Oruklu, A. Cinar, L. Quinn, and D. Smith, Estimation of Future Glucose Concentrations with Subject-Specific Recursive Linear Models, *Diabetes Technology & Therapeutics* **11** (2009), 243-253.
- [22] V. Marmarelis and G. Mitsis, *Data-driven Modeling for Diabetes: Diagnosis and Treatment*, Springer Science & Business, 2014.
- [23] C. Chatfield, Calculating Interval Forecasts, *Journal of Business & Economic Statistics* **11** (1993), 121-135.
- [24] B.E. Hansen, Interval forecasts and parameter uncertainty, *Journal of Econometrics* **135** (2006), 377-398.
- [25] Y.S. Lee and S. Scholtes, Empirical prediction intervals revisited, *International Journal of Forecasting* **30** (2014), 217-234.
- [26] T.C. Chan, C.C. King, M.Y. Yen, P.H. Chiang, C.S. Huang, and C.K. Hsiao, Probabilistic daily ILI syndromic surveillance with a spatio-temporal Bayesian hierarchical model, *PLoS One* **5** (2010), e11626.

- [27] L. Chen, K.S.M.T. Hossain, P. Butler, N. Ramakrishnan, and B.A. Prakash, Flu Gone Viral: Syndromic Surveillance of Flu on Twitter Using Temporal Topic Models, in: *2014 IEEE International Conference on Data Mining*, 2014, pp. 755-760.
- [28] C. Chung-Kuo, J.M. Overhage, and J. Huang, An application of sensor networks for syndromic surveillance, in: *Proceedings. 2005 IEEE Networking, Sensing and Control, 2005.*, 2005, pp. 191-196.
- [29] T.C. Weng, T.C. Chan, H.T. Lin, C.K. Chang, W.W. Wang, Z.R. Li, H.Y. Cheng, Y.R. Chu, A.W. Chiu, M.Y. Yen, and C.C. King, Early detection for cases of enterovirus- and influenza-like illness through a newly established school-based syndromic surveillance system in Taipei, January 2010 ~ August 2011, *PLoS One* **10** (2015), e0122865.
- [30] W. Yan, L. Palm, X. Lu, S. Nie, B. Xu, Q. Zhao, T. Tao, L. Cheng, L. Tan, H. Dong, and V.K. Diwan, ISS--an electronic syndromic surveillance system for infectious disease in rural China, *PLoS One* **8** (2013), e62749.

Address for correspondence

Ashenafi Zebene Woldaregay, Department of Computer Science,
University of Tromsø - The Arctic University of Norway, 9037
Tromsø, Norway, Email: ashenafi.z.woldaregay@uit.no

Uncovering Hidden Topics in Hong Kong Clinical Research Through Hospital Authority Convention Publications

Zoie Shui-Yee Wong^{a,b}, Wai-Man Chan^c, Eliza Lai-Yi Wong^d,
Patsy Yuen-Kwan Chau^d, Kwok-Leung Tsui^c, Hong Fung^d

^a Graduate School of Public Health, St. Luke's International University, Tokyo, Japan

^b School of Public Health and Community Medicine, The University of New South Wales, Sydney, Australia

^c Department of Systems Engineering and Engineering Management, City University of Hong Kong, Hong Kong, China

^d JC School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong, China

Abstract

Uncovering clinical research trends allows us to understand the direction of healthcare services and is essential for longer-term healthcare planning. The Hospital Authority Convention is a mainstream annual healthcare conference that gathers up-to-date Hong Kong medical research. We propose to use state-of-the-art medical document mining and topic modelling methods to uncover latent themes and structures in the publications. We collected 742 articles from HA Convention from the year 2013 to 2016 and selected 56 publications from the category of "Clinical Safety and Quality Service" for further analysis. Applying natural language processing and Latent Dirichlet Allocation (LDA) methods, we identified 7 potential topics, namely: surgical operation, hospital discharge, medical error, nursing procedure, service performance assessment, patient and staff engagement, and admission algorithm and standardisation. This exploratory study demonstrates that key themes exist in the annual HA Convention and we observe potential changes in healthcare services focus over the years in the selected category.

Keywords:

Data Mining; Unsupervised Machine Learning, Medical Informatics

Introduction

Uncovering clinical research foci and trends can allow healthcare providers, policy makers, researchers and general public to gain an overall picture on the direction of public health. For policy makers, through better understanding of health trends and services provided, new policies can be devised in order to meet the needs of ever-changing health market. For healthcare service providers, understanding health trends and shifts in focus provides good guidance on their current practice and may direct their future health services or business models. For the general public, gaining an appreciation of new innovations and trajectories in healthcare organisations allows them to exercise improved judgement in their consumption of health care products and services.

In Hong Kong, the Hospital Authority (HA) Convention is an annual conference and research exchange platform that gathers and shares up-to-date medical research among local public and

private hospitals, government departments, and local and overseas businesses. This convention acts as a parameter of development in the local healthcare system and quality improvement trends. This allows experience sharing and introduction of new healthcare-related services across inter-organisational levels to enhance the quality of Hong Kong healthcare services. Through an improved understanding of the active healthcare services provided and research carried out by different service providers, researcher and related parties, we may be able to better match between local demand and services. Therefore, the potential shifts on the focuses on the HA convention could give brief insights on the caused impacts.

Current research advances in natural language processing and topic modelling allow us to catalogue raw data into libraries for classification, aggregation, identification and document location. This can enable resource discovery and serves for archival and managerial purposes. The key challenge is that these documents are usually analysed in a large-scale manner and previous studies have revealed several unique problems in uncovering medical documents topics [1]. Discovering major themes from the corpus of data would facilitate retrieval of documents on the same topic with and without a query. We do not observe any similar health informatics studies in Hong Kong that aim to reveal the shifts of the focuses on healthcare sector, in particular on the angle on reviewing articles over the years using document text mining and topic modeling methods systematically.

We propose to apply state-of-the-art text mining and topic modelling techniques on medical documents to uncover latent themes and structure from HA Convention publications. In this probing study, we analyse metadata and abstracts from HA Convention publications, in particular the category of Clinical Safety and Quality Service research – that includes projects aim to (1) improve efficiency and effectiveness of care delivery/resources management and (2) enhance clinical safety and outcomes/clinical governance.

Methods

Unsupervised topic modelling

Topic modeling is a statistical tool that is frequently used in text mining for identifying common themes in a collection of documents [2]. This method retrieves information based on statistics of words in the documents and has been widely adopted in the

healthcare informatics discipline. One of its key potential applications is to identify health related topics from numerous articles. Wang et al. research on a typical topic modelling approach - Latent Dirichlet Allocation (LDA) model - to identify latent topics from the life science publications extracted from electronic databases [3]. Paul and Dredze [4] applied Ailment Topic Aspect Model (ATAM) and LDA models to identify coherent clusters with the Twitter data and demonstrated strong correlations with independent surveillance and survey data. Hashimoto et al. [5] incorporated topic model to identify keywords and support active learning in systematic reviews. Speier et al. [6] demonstrated that the application of topic modelling is useful for automatic summarisation system for clinical reports.

In this probing study, we applied a classical LDA model with the following procedure and assumptions [7-8]:

1. The generative process for a document $\omega = (\omega_1, \dots, \omega_N)$ of a corpus D containing N words from a vocabulary consisting of V different terms, $\omega_i \in \{1, \dots, V\}$ for all $i = 1, \dots, N$.
2. The term distribution of each topic, β is determined by $\beta \sim \text{Dirichlet}(\delta)$ and it contains the probability of a word occurring in a given topic.
3. The proportions θ of the topic distribution in a document, ω , are determined by $\theta \sim \text{Dirichlet}(\alpha)$.
4. For each of the N words ω_i
 - (i) A topic was chosen from the distribution, $z_i \sim \text{Multinomial}(\theta)$.
 - (ii) A word ω_i was chosen from a multinomial probability distribution conditioned on the topic $z_i: p(\omega_i | z_i, \beta)$

Variational expectation-maximisation (VEM) procedure has been adopted and its estimation on the log-likelihood for one document $\omega \in D$ is for LDA given by

$$\begin{aligned} \ell(\alpha, \beta) &= \log(p(\omega | \alpha, \beta)) \\ &= \log \int \left\{ \sum_z \left[\prod_{i=1}^N p(\omega_i | z_i, \beta) p(z_i | \theta) \right] \right\} p(\theta | \alpha) d\theta \end{aligned}$$

In comparison to other black-box algorithms, the strength of topic model is that it interprets clusters by the word probability distributions over topics. Furthermore, it allows terms associated with a mixture of topics rather than only one single topic. This characteristic was regarded as a key advantage for various real life applications, including bioinformatics [9].

Metadata, modelling procedure and underlying assumptions

All the contributions from the HA Convention offered metadata information and context of the research that included year of publication, session, authors and their corresponding affiliations, research topic, abstracts and keywords. The data was extracted from the online HA Convention programme books published from 2013 to 2016.

According to medical text structuring method conducted by Wong [10], we pre-processed the textual medical publications into structured data output (i.e. bag of word format). LDA is one of the typical topic models and is an unsupervised generative graphical model [2, 11]. We assumed the potential latent topics were uncorrelated and documents were considered as

random mixtures over latent topics, which each topic is characterised by a distribution of various keywords. Those keywords are generated via repeated sampling on a topic based on the topic distribution and selection of a word given the chosen topics.

In order to identify the latent topics among these abstracts without specifying the keywords, an unsupervised topic modelling method (i.e. the LDA model with the VEM procedure for maximum likelihood estimation and Gibbs sampling method) was applied. Topics and the representing keywords were then discovered. The entire analysis was conducted on a R v3.3.2 (64-bit) platform using 'NLP', 'tm', 'tau', 'Rstem', 'SnowballC' and 'topicmodels' [12] packages. The modelling procedure is shown in Figure 1.

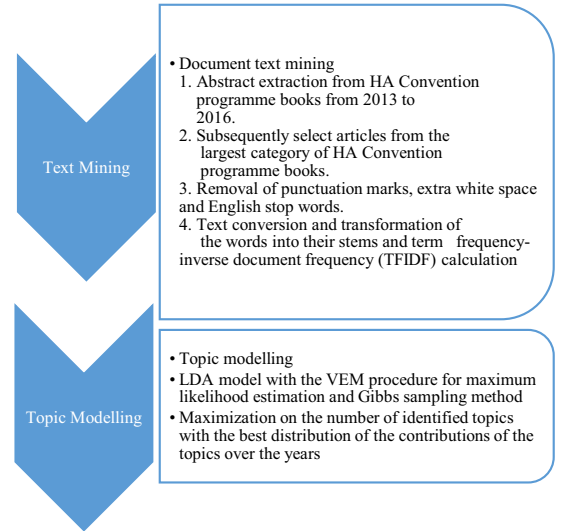


Figure 1– Data preparation and modelling procedure

Results

In this section we arrange our findings into two parts. The first part presents descriptive statistics of metadata information and summarises document data mining outcomes and the second part shows the uncovered topic structure, including per-document topic distributions, the per-document per-word topic assignments, and the corresponding potential topics/research themes.

Descriptive statistics

We collected 742 articles from the HA Convention from the years 2013 to 2016. The numbers of participating organisations and articles increased gradually from 2013 to 2016. In 2016, there were 229 included articles from 84 organisations in the HA Convention. Most articles were categorised under the sessions of “Special Topics” and “Masterclasses” while the “Service Enhancement Presentations” was the session with the largest number of publications over the 4 years, as shown in Table 1. In total, 2,087 authors from 921 affiliations have submitted their works to the Convention. Twelve % of overall affiliations from local hospitals were from New Territories East cluster and 16% were from local hospitals in Kowloon Central cluster.

Table 1- Number of publications (%) in different sessions in years

Sessions	HA Convention			
	2013	2014	2015	2016
Plenary Sessions	11 (7.86%)	9 (4.92%)	8 (4.21%)	10 (4.37%)
Parallel sessions	16 (11.43%)	15 (8.20%)	16 (8.42%)	0 (0.00%)
Symposiums	27 (19.29%)	42 (22.95%)	30 (15.79%)	28 (12.23%)
Special Topics	8 (5.71%)	25 (13.66%)	29 (15.26%)	59 (25.76%)
Masterclasses	13 (9.29%)	21 (11.48%)	34 (17.89%)	52 (22.71%)
Service Enhancement Presentations	56 (40.00%)	57 (31.15%)	56 (29.47%)	56 (24.45%)
Corporate Scholarship Presentations	9 (6.43%)	14 (7.65%)	16 (8.42%)	16 (6.99%)
Others	0 (0.00%)	0 (0.00%)	1 (0.53%)	8 (3.49%)
# of articles	140	183	190	229
*# of contributed organisations	68	68	81	84

The convention has several research categories, for instance: “Clinical Safety and Quality Service / Quality and Safety in Healthcare”, “eHealth and Healthcare”, “Palliative Care”, “Allied Health Service – Technology and Innovation”, and “Committed and Happy Staff”. Among these, “Clinical Safety and Quality Service / Quality and Safety in Healthcare” is the largest category with the highest number of contributions over the 4 years - which accounts for about 8% of the all publications. Fifty-six articles under this category were extracted for further natural language processing and topic model analysis. After removing stop words and punctuation and applying stemmer, 2,200 terms were identified and these were the inputs to build a document matrix. The top 10 TFIDF stemmed terms are reported as “drug”, “fall”, “count”, “vap”, “consent”, “specimen”, “nurs”, “admiss”, “duti”, and “group”.

Table 2- Selected keywords of the seven identified topics under the category of “Quality and Safety in Healthcare / Clinical Safety and Quality Service” from 2013 to 2016

Seven identified topics and the included corresponding keywords						
1	2	3	4	5	6	7
tavi	hospit	treat- ment	inform	surgic	patient	risk
oper	discharg	cathe	record	perform	staff	admiss
post review	fall safeti	error healthca r	imv speci- men	strategi audit	nurs drug	score wound
audit	consent	oper	train	emerg	pro- gramm servic	stand- ardis perform
nurs	pci	phar- maci mrsa	vre	surgeon		
vap	bed		work- flow environ- ment	monitor	imple- ment qualiti	assign clinic
check	incid	emerg		recom- mend		
coordin	surgeri	haemo- dialysi	intak	ap- proach	team	health
curtain	vap	cpb	medicin	check	inter- vent	inpati

Topic model

We developed LDA model with VEM procedure for maximum likelihood estimation and Gibbs sampling method through Statistics R. The document matrix summary indicates sparsity of 54% - which is lower than the threshold value of 95%. We did

not remove words from the matrix for the subsequent construction of topic modelling and we adjusted the number of topics manually by trial and error in order to maximise the number of the topics. Seven topics were identified and Table 2 shows the representative keywords of each identified topic.

These findings were further discussed and verified with local Public Health and Health System experts in Hong Kong. Based on the top keywords of each topic, we suggested that the possible seven themes of topics under ‘Quality and Safety in Healthcare’/ ‘Clinical Safety and Quality Service’ session are: (1) surgical operation, (2) hospital discharge, (3) medical error, (4) nursing procedure, (5) service performance assessment, (6) patient and staff engagement, and (7) admission algorithm and standardisation, as shown in Table 3.

Table 3- Potential themes of the identified topics

Topic	Potential theme
1	Surgical operation
2	Hospital discharge
3	Medical error
4	Nursing procedure
5	Service performance assessment
6	Patient and staff engagement
7	Admission algorithm and standardisation

* Study preliminary finding requires further investigation

We summarised the per-document topic distributions annually using the identified topics and Figure 2 shows the annual document topic distributions for each topics under the session of “Quality and Safety in Healthcare” / “Clinical Safety and Quality Service”. This result indicates that that each of the Topics 1 to 4 (i.e. potentially surgical operation, hospital discharge, medical error, nursing procedure) offers more than 30% contribution to a corresponding year - which can be potentially considered as the key topics in those years. Topic 6 (potentially patient and staff engagement) is considered as the common topics found over the 4-year programmes as the keywords show similar and significant contributions over years, while Topic 5 (potentially service performance assessment) can be catalogued as minor topic of this theme in 2014 HA programme book. Topic 7 (potentially admission algorithm and standardisation) is considered as another minor topic covered in year 2014 to 2016.

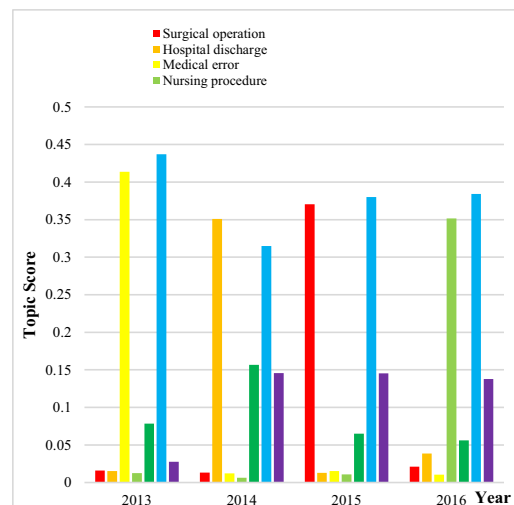


Figure 2 - Contributions of the seven identified topics under the category of ‘Quality and Safety in Healthcare’/ ‘Clinical Safety and Quality Service’ from 2013 to 2016.

Discussion

We selected 56 articles from HA Convention programme and incorporated document text mining and topic modelling to uncover 7 potential topics for the category of “Quality and Safety in Healthcare”/ “Clinical Safety and Quality Service”. The possible topics include: “Surgical operation”, “Hospital discharge”, “Medical error”, “Nursing procedure”, “Service performance assessment”, “Patient and staff engagement”, and “Admission algorithm and standardisation”. The finding suggests that there may have at least one dominant key topic for each year’s publication, which implies that in the context of enhancement on quality and safety of the clinical services provided, the focus of research topics may potentially be different annually. Some topics’ contribution distributed relatively equal among years and this implies those are relatively long term and ongoing research issues over several years. These findings fulfill one of the major objectives on uncovering the potential trends and shifts on health research studies in Hong Kong via the use of informatics and modelling techniques.

With a better catalogued theme, we can analyse the development (and their causes) in the number and proportion of clinical studies over time in Hong Kong that projects the need of the healthcare. By analysing differences in trends, future study can be designed to demonstrate the cause of different trajectories over years and to answer some policy research questions, such as clinical development continues to evolve against any pattern of increasing development timelines and associated costs. In addition, it is also believed that these publication trends could reflect how the healthcare providers responded to the implementation of new policies. In this explorative study, we demonstrated that key focusing topics exist in the abstracts from presentations of annual HA Convention and potential changes on the focus of the healthcare services have been identified over the years. We foresee that the local research trend may potentially point to topic on safeguarding consumers in the future while the planned accreditation professional scheme for allied health professionals was raised by the Hong Kong government recently in 2016.

Although we have identified the latent topics in the category of “Quality and Safety in Healthcare / Clinical Safety and Quality Service”, there are several potential research directions for our further investigation. Currently, we cannot be certain if the interpretation of the identified topics is absolute. It is necessary to scrutinise this results in a careful manner – not just evaluate the outcome on the basis of the keywords found. It appears that the keywords found under each topic are relatively diversified, which make the key focused themes unclear through pure inspection over the top key terms. To improve this study, we plan to validate this preliminary finding via other complementary unsupervised learning methods. Furthermore, we will carry out in-depth discussion with further hospital practitioners and experts in healthcare quality and safety to interpret the public health and clinical meaning of the model outcomes. The modelling and knowledge extraction procedure can be rectified to improve the representativeness of the mined outcomes [13]. For instance, some of the identified keywords with little or no clinical relevance could be filtered out under the consensus of 2 investigators.

An improved design text extraction procedure may be necessary to effectively select potential text variables for topic modelling. For instance, further studies can be carried out to incorporate other state-of-the-art natural language processing techniques, such as ‘n-gram’ and ‘wordnet’ to improve the accuracy on the identification and interpretation of the topics.

“n-gram” is a package that commonly used in document text mining to identify co-occurring words within a given window while synonyms can be identified with the use of the package of “wordnet”. In addition, term associations can be tested in order to provide more comprehensive meanings on the identified keywords. This systematic approach can be designed in order to avoid introduction of selection bias through training-up the model.

Another future direction could be to compare the model outcomes using unsupervised machine learning and supervised machine learning. Since the LDA model is an unsupervised machine learning method and its accuracy highly depends on the occurrence of the keywords with representatives on specific topics. For example, the authors may not include the words “blood pressure” in an abstract on the evaluation of the performance on newly developed drugs on handling hypertensive patients, but the article could fall under the sub-category of cardiology. Therefore, it may be useful to explore and evaluate the strengths of various machine learning and topic modelling approaches and incorporate a suitable modelling procedure to better reveal the potential nature of hidden topics.

The “topics” discovered in an unsupervised manner may not always genuinely reflect the true topics from the data. Also, the classical LDA method instinctually does not consider the order of words in a document. Apart from the classic unsupervised topic discovery we explored in this study (that do not involve any prior annotations or labeling of the documents), we suggest to investigate the possibility to use other supervised learning methods, such as supervised LDA (sLDA), and incorporate topic hierarchies from data via hierarchical LDA (hLDA) method for further analysis. The performance of various methods can be evaluated by experiments with text datasets of medical documents. The model performance can also be evaluated quantitatively through log-likelihood on held-out data approach. We will also collect expert views and independent health research indicators and materials for validating the model outcome. A better structured analytical framework is potentially useful for analysing other global clinical research papers from mainstream clinical databases, such as PubMed and Medline.

Conclusion

We used the HA Convention publications to infer the hidden topic structure in healthcare quality and safety research in Hong Kong. This explorative study provides opportunities to reveal the major focus and trends on the healthcare services provided by different organisations from year 2013 to 2016. This research contributes to the emerging field of understanding the characteristics of the medical documents and how to account for them using health informatics methods. This work potentially enhances our ability to interpret medical publication information and trends in Hong Kong and globally.

Acknowledgements

This research was supported by the Research Grants Council Theme-Based Research Scheme (Ref.: T32-102/14N) and Collaborative Research Fund (Ref. CityU8/CRF/12G).

References

- [1] A.G. Karami, Aryya; Zhou, Bin; Kharrazi, Hadi, A Fuzzy Approach Model for Uncovering Hidden Latent Semantic Structure in Medical Text Collections, in: *iConference 2015* 2015.

- [2] D.M. Blei, Probabilistic topic models, *Commun. ACM* 55 (2012), 77-84.
- [3] H. Wang, Y. Ding, J. Tang, X. Dong, B. He, J. Qiu, and D.J. Wild, Finding Complex Biological Relationships in Recent PubMed Articles Using Bio-LDA, *PLoS One* 6 (2011), e17243.
- [4] M.J. Paul and M. Dredze, Discovering Health Topics in Social Media Using Topic Models, *PLoS One* 9 (2014), e103408.
- [5] K. Hashimoto, G. Kontonatsios, M. Miwa, and S. Ananiadou, Topic detection using paragraph vectors to support active learning in systematic reviews, *J Biomed Inform* 62 (2016), 59-65.
- [6] W. Speier, M.K. Ong, and C.W. Arnold, Using phrases and document metadata to improve topic modeling of clinical reports, *J Biomed Inform* 61 (2016), 260-266.
- [7] Grun, B and Hornik K, topicmodels: An R Package for Fitting Topic Models, available at <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf>
- [8] J. Chang, lda: Collapsed Gibbs Sampling Methods for Topic Models, <https://cran.r-project.org/web/packages/lda/lda.pdf>
- [9] L. Liu, L. Tang, W. Dong, S. Yao, and W. Zhou, An overview of topic modeling and its current applications in bioinformatics, *SpringerPlus* 5 (2016), 1608.
- [10] Z.S. Wong, Statistical classification of drug incidents due to look-alike sound-alike mix-ups, *Health Informatics J* (2014).
- [11] D.M. Blei, A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research* 3 (2003), 993-1022.
- [12] Grun, B and Hornik K, available at <https://cran.r-project.org/web/packages/topicmodels/topicmodels.pdf>
- [13] G. Karystianis, T. Sheppard, W.G. Dixon, and G. Nenadic, Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database, *BMC Med Inform Decis Mak* 16 (2016), 18.

Address for correspondence

Corresponding author: Dr. Zoie Shui-Yee Wong. Address: Graduate School of Public Health, St. Luke's International University, 5/F, 3-6-2 Tsukiji, Chuo-ku, Tokyo, 104-0045 Japan.

Aligned-Layer Text Search in Clinical Notes

Stephen Wu^a, Andrew Wen^a, Yanshan Wang^b,
Sijia Liu^b, Hongfang Liu^b

^a Department of Medical Informatics and Clinical Epidemiology,
Oregon Health & Science University, Portland, OR, USA,

^b Division of Biomedical Informatics, Mayo Clinic, Rochester, MN, USA

Abstract

Search techniques in clinical text need to make fine-grained semantic distinctions, since medical terms may be negated, about someone other than the patient, or at some time other than the present. While natural language processing (NLP) approaches address these fine-grained distinctions, a task like patient cohort identification from electronic health records (EHRs) simultaneously requires a much more coarse-grained combination of evidence from the text and structured data of each patient's health records. We thus introduce aligned-layer language models, a novel approach to information retrieval (IR) that incorporates the output of other NLP systems. We show that this framework is able to represent standard IR queries, formulate previously impossible multi-layered queries, and customize the desired degree of linguistic granularity.

Keywords:

Natural Language Processing; Information Storage and Retrieval; Electronic Health Records

Introduction

Search (i.e., information retrieval) techniques in clinical text face the challenge of coarse vs. fine granularity. In a setting such as electronic health records (EHRs), the goal is usually to find a broad characterization of a whole patient from that patient's diverse assemblage of text and structured data. How to combine these pieces of evidence into a patient-level judgment is an area of active research. Moreover, in clinical text, dictionary matching for a symptom like "hypertension" must be augmented by fine-grained algorithms to ensure that the symptom is not negated (e.g., "no complaints of ..."), about someone other than the patient (e.g., "family history of"), or a host of other possibilities. Textual variants of these fine-grained patterns abound as well, spawning natural language processing (NLP) methods to find concepts of interest from acceptable textual contexts.

Therefore, we aim to incorporate arbitrary NLP-derived features into information retrieval (IR) methods on clinical text, allowing for greater control of the granularity in search. Our main contribution is the *aligned-layer information retrieval* model, which we specify by defining the "layered" nature of text and the nature of phrases in this setting. This model is a straightforward extension of the language modeling approach to IR and its feature-centric successors.

We adopt the "layer" terminology from the language resources community in which, for example, treebanking and propbanking are different layers of annotations on the same text. The crucial contribution in our work is a model in which all text-derived "layers" are aligned by token, and can be utilized and scored simultaneously during search.

After describing our aligned layer approach, we report on a preliminary system implementation and evaluation of the model. For our evaluation, we focus on the problem of cohort identification within clinical notes from the EHR.

Related Work

Preliminary explorations with concepts (from the Unified Medical Language System (UMLS)) in language modeling have met with moderate success [1-3] in the past. Language modeling in IR [4-6] builds on a rich tradition of probabilistic IR [7], and has a successful history of ranking documents based on well-motivated textual features. The most commonly-used textual features are from Metzler and Croft's dependence model, modeling the probabilistic relationships between query terms (and a candidate document) as Markov Random Fields (MRFs). These term-term dependencies significantly outperformed the original bag-of-words language modeling approaches to IR from which they arose [4; 5]. Subsequently, query hypergraphs made it possible to model higher-order dependencies (e.g., multiword-multiword or multiword-term dependencies) [8]. While the representational power of these feature-based models allows for arbitrary *non-textual* features as well (e.g., named entities or dependencies, such as we introduce here), non-textual features have received minimal attention in the literature.

Recent work on joint text and concept search [2] and split-layer language models [1] extended the notion of mixing different document representations together [9; 10]. Arising out of the medical domain, these techniques all showed some benefit to considering multiple "layers" simultaneously. Our work here extends the discussion of language modeling layers, pushing the question of what semantic representations (or granularity thereof) are effective for IR.

Positional language models [11] attempt to model the intuition that closer textual proximity can correspond to greater association. This shares an important intuition with aligned-layer models: position is important in IR. However, our accounting of position is not to model the textual distance, but the content and relationships that are latent within text at the same locations. A number of existing studies [1; 2] require an NLP preprocessing step prior to indexing the collection, and utilize these text-derived features in search. However, these attempts fail to allow navigation among, and correspondences between, those structures.

The task of query formulation [12] is somewhat upstream task to our layered models; query formulation can include the introduction of term dependencies [13], weighting of terms [14], query expansion [15], and the parameterization of associated weights according to additional corpora [16; 17]. Our hope is that future query formulation techniques will make use of aligned-layer structures.

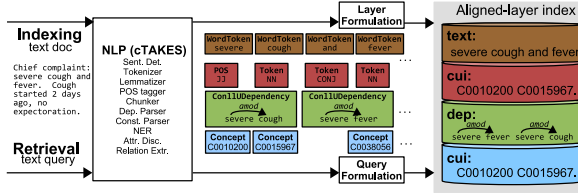


Figure 1 - cTAKES processing is followed by the indexing of results into various layers (indexes)

The TREC Medical Records track [18; 19] provided resources for significant innovation in medical IR, which used in this paper for evaluation: collections of clinical text, information

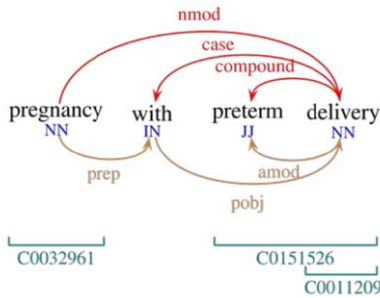


Figure 2 - Aligned layers (tokens, part-of-speech tags, named entities, Stanford dependencies, and CoNLL-X dependencies) for a sample query “pregnancy with preterm delivery.” Span, source, and target indices are with respect to the L_0 layer’s indices.

needs, and judgments. Language modeling techniques have been successful in this clinical text setting [2], as well as in health-related web search exemplified by the CLEF eHealth challenges [20].

Methods

We now turn our attention to the aligned-layer retrieval model and its implementation for our experiments. Let us assume that we run an NLP pipeline as a preprocess; e.g., the clinical Text Analysis and Knowledge Extraction System (cTAKES)¹ [21]. Developed specifically for clinical text, this pipeline produces NLP artifacts as illustrated in the left and middle Figure 1.

Indexing: Aligning text and NLP artifacts

We represent this NLP-preprocessed text as a finite sequence of multiple aligned layers. Specifically, a document or query is composed of multiple layers L_0, L_1, L_2, \dots . We reserve L_0 to be a *base layer* of the original text; namely, a sequence of tokens (optionally stopped and stemmed) by which all other layers are aligned.

Figure 2 illustrates a query text (tokens) as a base layer, with several underlying layers—part-of-speech (POS) tags, named entities (NEs) with mappings to concept unique identifiers (CUIs), and two types of dependency parses (left). It then shows their translation into aligned layers (right).

Other than the base layer, each layer is an *artifact layer*, and is composed of a sequence of *artifacts*. For example, in Figure 2, L_2 is named entities, and $L_2 = l_{2,0}, l_{2,1}, l_{2,2}, \dots = C0032961, C0151526, C0011209, \dots$. The artifacts are aligned with the base layer by storing 2 additional numbers:

start index and length. Thus, a function $span(\cdot)$ on $l_{2,1}$, a named entity spanning the words “preterm delivery,” would have $span(l_{2,1}) = (2,2)$ since the artifact concerns “preterm delivery” and starts at index 2 with a length of 2 tokens.

Additionally, we define a special *relation* artifact with a slight modification of other artifacts. For relation layers like L_3 (stanford dependency parses, in Figure 2), artifacts are relations between two other artifacts’ locations. We will write these as $L_3 = r_{3,0}, r_{3,1}, r_{3,2}, \dots$ to emphasize that these artifacts are relations. Then, $source(\cdot)$ and $target(\cdot)$ give the position and length of each of the relation’s arguments; $source(r_{3,0}) = (0,1)$ points from “pregnancy” to $target(r_{3,0}) = base(r_{3,0}) = (3,1)$ “delivery.”

text tok	$l_{0,0}$ pregnancy	$l_{0,1}$ with	$l_{0,2}$ preterm	$l_{0,3}$ delivery
pos tag	$l_{1,0}$ span=(0,1) NN	$l_{1,1}$ span=(1,1) IN	$l_{1,2}$ span=(2,1) JJ	$l_{1,3}$ span=(3,1) NN
named ent	$l_{2,0}$ span=(0,1) C0032961	$l_{2,1}$ span=(2,2) C0151526	$l_{2,2}$ span=(3,1) C0011209	
stanf dep	$r_{3,0}$ source=(0,1) target=(3,1) nmod	$r_{3,1}$ source=(3,1) target=(1,1) case	$r_{3,2}$ source=(3,1) target=(2,1) compound	
conll dep	$r_{4,0}$ source=(0,1) target=(1,1) prep	$r_{4,1}$ source=(1,1) target=(3,1) pobj	$r_{4,2}$ source=(3,1) target=(2,1) amod	

Note that the transformation to aligned language layers is slightly lossy. In Figure 2 the L_3 artifacts corresponding to the head and dependent of each dependency relation are not preserved—only the L_0 -aligned spans. The loss is minimal in most NLP structures of interest.

Scoring: Terms and phrases

In IR language models, it is common to rank according to $score(D, Q) = P(D) \cdot P(Q|D)$. The conditional probability encapsulates the intuition that an ad hoc user trying to find document D will try to write an effective query Q . We focus our attention on the conditional probability $P(Q|D)$.

This is most simply expressed in the standard *query likelihood* model with Dirichlet smoothing (with the parameter μ_D), a baseline for language modeling approaches. We write *single-term queries* with the notation of the aligned layers above, where the a variable in l_a simply indicates that such term queries can be written for any single layer:

$$p_t(l_a|D) = \frac{\#t_D(l_a) + \mu \frac{\#t_D(l_a)}{|D|_a}}{|D|_a + \mu} \quad (1)$$

where $\#t_D$ is the number of occurrences of the argument (a query term l_a) in the document; $\#t_D$ is the number of occurrences of that argument in the whole collection. Similarly, $|D|_a$ is the number of artifacts in the document from layer l_a ; $|D|_a$ counts the same layer’s artifacts in the whole collection.

With a layered representation of both queries and documents, these term operators allow for querying of artifacts in any layer L_a . The standard query likelihood model in other texts [5] is then just a special case of our term query utilizing only text layer artifacts L_0 (i.e., from L_{text}). The smoothed

¹ <http://ctakes.apache.org>

probability is also calculated with respect to frequencies in layer L_a . Term queries do not require *aligned* layers — they consider each layer separately.

Probabilities for individual query terms can be calculated separately and combined to produce an overall probability. This is done implicitly, but can be made explicit with a **list query**; items in the list need not be in the same layer because they are each considered individually. A list query also groups the items in the list for use in other queries.

However, list queries ignore the collocation of layered artifacts; thus, we define the **phrase query**. A phrase is two or more artifacts from an arbitrary combination of layers, e.g., $l_{ai}l_{bj}$, the i^{th} artifact in L_a and the j^{th} artifact in L_b . Phrases must be specified with ordering o as ordered or unordered (True or False), with a window size w to search within (measured according to positions in the base layer L_0).

$$P_{\text{ph}}(l_{ai}l_{bj} \dots | D; o, w) = \frac{\# \text{ph}_D(l_{ai}l_{bj} \dots) + \mu \frac{\# \text{ph}_D(l_{ai}l_{bj} \dots)}{|D|_a}}{\max(|D|_a, |D|_b, \dots) + \mu} \quad (2)$$

The function $\# \text{ph}_D$ counts the number of cross-layer phrase matches in a document. Having aligned $l_{ai}l_{bj} \dots$ to the base layer L_0 , we can check across layers for matches within a window. Window length is between the end of the first matched artifact, and the beginning of the last matched artifact (i.e., the last artifact only needs to have start index within the window length; it can end outside the window). Writing $l_{ai}l_{bj}$ illustrates that artifacts need not arise from the same layer, and will be in different positions within their respective layers (though aligned to the base layer L_0). However, ordered phrases with $o=\text{True}$ must additionally consist of only non-overlapping, sequential artifacts.

Whereas in single-layer queries, a probability estimate would consist of all items in that layer, there are now two or more layers to consider. It may at first seem that the denominator should be the product $|D|_a \cdot |D|_b \cdot \dots$. This would mean: “out of all possible multi-layered phrases that could be constructed in document D , how likely is the specific construction we are looking for?” This choice of denominator would yield sparse probability estimates. Instead, we have chosen the size of the largest layer stipulated in the phrase. This answers the question: “of the artifacts in the largest layer, what proportion participate in a phrase consistent with the query?” This may be thought of as a backed-off estimate, and is more tractable. The choice of a denominator is an interesting area of further investigation beyond the scope of this work.

We should also note that this aligned-layer phrase queries subsume ordered and unordered phrases of feature-centric IR models [6], since those features are simply functions of L_0 .

Implementation

The fundamental aims of the aligned-layer IR model require significant extension in any search engine. We implemented an aligned-layer language model via a plugin to Elasticsearch with multiple components, tested with Elasticsearch 1.7.2. We used cTAKES v3.2 as a preprocess [22], producing *character-aligned* artifacts (based on Apache UIMA²). Aligned-layer models are similar to searching over these UIMA-based data structures, but with scored ranking, across multiple documents, and aligned on tokens instead of characters.

For indexing, each layer was represented as a field within a document. For our evaluation and analysis, we generated the an index with the following fields:

- L_{text} : Word tokens, as identified by the cTAKES tokenizer; this is considered the base layer L_0
- L_{cui} : Concept Unique Identifiers (CUIs), as mapped from the UMLS Metathesaurus by cTAKES
- L_{tui} : Type Unique Identifiers, a many-to-one mapping that groups concepts (CUIs) into semantic types (TUIs)
- L_{lemma} : Normalized by a version of the National Library of Medicine's (NLM) Lexical Variant Generator (LVG)
- L_{dep} : Conll-U dependency parses from Clear Parser [23] trained on treebanked clinical text from Mayo Clinic
- L_{pos} : Part-of-speech tags produced by the OpenNLP³ MaxEnt POS tagger, trained on treebanked clinical text from Mayo Clinic

Each artifact was marked with its span (position and length, in tokens) via Lucene Payloads; this is a trivial marking for the base layer of text L_0 , but can be significant in other layers, such as for multi-word expressions in the CUI layer of Figures 1 and 2. In addition, the size of each field within a document was stored as metadata with that document at indexing time.

Queries are scored in Elasticsearch and Lucene via a highly optimized scoring interface. However, our scoring functions cannot be represented within that original structure; for example, the Lucene implementation of Dirichlet smoothing on a language model fails to divide by the collection frequency in some cases. Therefore, we implemented scoring via Elasticsearch Script Queries, which are typically used in a filtering context but here provide us the flexibility to score according to the models defined above.

Because they make use of aligned-layer terms and phrases, queries must be constructed and parsed differently than in other search systems. We implemented our own query parser in JavaCC to term and phrase operators in the aligned-layer query language.

Task and Experiments

We provide a preliminary evaluation of aligned-layer language models, and of our Elasticsearch-based implementation in particular, on the task of patient cohort identification as exemplified in the 2011-2012 Text Retrieval Conference (TREC) Medical Records Track, or TREC-Med [18; 19]. In brief, there were 81 topics (34 from 2011, 47 from 2012) such as “patients with hearing loss,” and the task was to return lists of relevant patient visits (a de-identified surrogate for whole patient records) from among 17,198 possible visits. System-produced lists of visits were compared with human relevance judgments (which were gathered for the TREC-Med 2011-2012 competitions).

Table 1 shows how the same query is represented across our evaluated approaches. As a baseline test TXT, we used unstemmed, stopped TREC-Med queries with the standard query likelihood model — namely, term queries on L_{text} . A second baseline was CUI, where we used term queries on L_{cui} .

In aligned-layer IR, term operators on multiple layers can be weighed together simultaneously for results. We replaced all named entity mentions in the query with the first-matched CUI, mixing the two layers in MIX. Since the cTAKES Named Entity Recognition (NER) module actually returned a set of CUIs per named entity, we included a model CUI-LS in which named entities were replaced with all of the CUIs associated with the same span, combined in a list. Next, we augmented

² <http://uima.apache.org>

³ <https://opennlp.apache.org>

the CUI list with a #ph containing L_{text} terms corresponding to any named entities, terming this PH-LS.

Table 1 - Eight versions of aligned-layer queries for Topic 125. #p are phrase queries, #l are list queries, others are term queries.

Layers	Query representation
TXT	coinfected hepatitis c hiv
CUI	cui:C0019196 cui:C0019158
MIX	coinfected cui:C0019158 hiv
PH-LS	coinfected #l(cui:C0019158 hepatitis) hiv
TXT-MRF	coinfected^0.85 hepatitis^0.85 c^0.85 hiv^0.85 #p(8 true coinfected hepatitis)^0.1 #p(8 false coinfected hepatitis)^0.05 #p(8 true hepatitis c)^0.1 #p(8 false hepatitis c)^0.05 #p(8 true c hiv)^0.1 #p(8 false c hiv)^0.05 #p(12 true coinfected hepatitis c)^0.1 #p(12 false coinfected hepatitis c)^0.05 #p(12 true hepatitis c hiv)^0.1 #p(12 false hepatitis c hiv)^0.05 #p(16 true coinfected hepatitis c hiv)^0.1 #p(16 false coinfected hepatitis c hiv)^0.05
CUI-MRF	cui:C0019196^0.85 cui:C0019158^0.85 #p(8 true cui:C0019196 cui:C0019158)^0.1#p(8 false cui:C0019196 cui:C0019158)^0.05
MIX-MRF	coinfected^0.85 cui:C0019158^0.85 hiv^0.85 #p(8 true coinfected cui:C0019158)^0.1 #p(8 false coinfected cui:C0019158)^0.05 #p(8 true cui:C0019158 hiv)^0.1 #p(8 false cui:C0019158 hiv)^0.05 #p(12 true coinfected cui:C0019158 hiv)^0.1 #p(12 false coinfected cui:C0019158 hiv)^0.05
PH-LS-MRF	coinfected^0.85 #l(cui:C0019158 hepatitis)^0.85 hiv^0.85 #p(8 true coinfected #l(cui:C0019158 hepatitis))^0.1 #p(8 false coinfected #l(cui:C0019158 hepatitis))^0.05 #p(8 true #l(cui:C0019158 hepatitis) hiv)^0.1 #p(8 false #l(cui:C0019158 hepatitis) hiv)^0.05 #p(12 true coinfected #l(cui:C0019158 hepatitis) hiv)^0.1

Searching for an optimal formulation and weighting of queries is outside the scope of this paper. However, for each of the above approaches, we also implemented a version with the Markov Random Field (MRF) “term” dependence model [24], which attempts to move beyond bag-of-words (BOW) models. Comparing the TXT-MRF row to the TXT row in Table 1 illustrates the behavior of this model: it tries out, with appropriate weights, the possibility that a multi-layered query may have phrases within it. We implemented a sequential dependence model with a limit of 5 aligned-layer components. These “term” dependence versions are marked as -MRF.

Results

We used mean average precision (MAP) as the primary evaluation measure for evaluation. MAP provides a single-figure measure of quality across recall levels [25]. While TREC-Med 2011 reported bpref [26] as its primary evaluation metric, and TREC-Med 2012 reported infAP [27], we here report the mean average precision (MAP) due to its stability for both training and testing in previous work [28; 29]. Table 2 shows the MAP scores, where we have separated the topics from TREC-Med 2011 (left) and 2012 (right).

Table 2 - Retrieval performance for a range of possible aligned-layer models

Model	2011		2012	
	BOW	MRF	BOW	MRF
TXT	0.2960	0.2936	0.2152	0.2224
CUI	0.3042	0.3119	0.3126	0.3101
MIX	0.2807	0.2847	0.2215	0.2167
CUI-LS	0.3076	0.3095	0.2161	0.2120
PH-LS	0.3185	0.3167	0.2172	0.2126

While this performance is below the state-of-the-art, it is interesting to note that it is layers with CUIs that obtain the highest performance in these tests (CUI-BOW on 2011 queries, and PH-LS-BOW on 2012 queries). Without text-layer query expansion, query logs, clickthrough data, or the like, another well-motivated, semantically-rich layer seems to be beneficial.

Discussion

For aligned-layer language models, it is interesting to ask whether there is the possibility of fine-tuning results on individual queries. Here, then, we consider one of the worst-performing queries, topic 121, and seek to write aligned-layer

queries that would improve those particular topics.

Topic 121 is “Patients with CAD who presented to the Emergency Department with Acute Coronary Syndrome and were given Plavix.” The CUIs found were C0948089 (acute coronary syndrome), C0010068 (coronary heart disease), and C0039082 (syndrome) – note the NER did not find the medication Plavix, or its generic form, clopidogrel. The highest scoring model on this topic was MIX-BOW=0.0639. In the event of a failed detection, we may consider using some kind of backoff model to less-specific expressions.

- In place of plavix, we searched for a dependency relationship where the head word has a lemma “give,” and the child is any word with a proper noun. It then weights that relation alongside the term plavix. This brought the MAP to 0.0803.
- In place of presented emergency department, we considered any verbs that might describe a patient's arrival. We stipulated that there should be some verb in a dependency relationship with the phrase “emergency department.” This brought the MAP to 0.0867.

The multi-layered queries clearly allow for greater coarseness or fineness according to the needs of a query.

Conclusion

We have introduced aligned-layer language models, a novel approach to IR that incorporates the output of other NLP systems. Core to this contribution are the definition of layers, alignment, and multi-layer scoring models. We have shown that this system can represent standard IR queries, as well as formulate multi-layered queries that were previously impossible. A case study demonstrates how the aligned-layer approach may feasibly be further extended to customize linguistic granularity to specific queries.

An open question of the proposed approach is how to design the layers for different corpora. Since different corpora contain variegated domain knowledge, we could design individualized layers for each corpus. In future work, we will further explore how our aligned-layer index can serve as a

corpus analysis tool for quantitative representativeness of linguistic features; this would further drive a model to automatically design and select layers. In addition to this manual approach to exploring the linguistic content, we will learn parameters and weights for the model to optimize performance on IR tasks. Finally, we will release the code for aligned-layer language models to the open-source community.

Acknowledgements

This work is funded in part by the U.S. National Institutes of Health grant R01LM011934. Thanks to Dingcheng Li, who was involved in early work on layered language models.

References

- [1] S. Wu, D. Zhu, W. Hersh, and H. Liu, Clinical Information Retrieval with Split-layer Language Models, in: *Proceedings of the SIGIR Workshop on Health Search and Discovery*, 2013.
- [2] D. Zhu and B. Carterette, Joint search in text and concept spaces for EMR-based cohort identification, in: *Bioinformatics and Biomedicine (BIBM)*, 2013 IEEE International Conference on, 2013, pp. 597-601.
- [3] S. Choi, J. Choi, S. Yoo, H. Kim, and Y. Lee, Semantic concept-enriched dependence model for medical information retrieval, *Journal of Biomedical Informatics* **47** (2014), 18-27.
- [4] J.M. Ponte and W.B. Croft, A language modeling approach to information retrieval, in: *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 275-281.
- [5] C. Zhai and J. Lafferty, A study of smoothing methods for language models applied to ad hoc information retrieval, in: *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2001, pp. 334-342.
- [6] D. Metzler, *A feature-centric view of information retrieval*, Springer Science & Business Media, 2011.
- [7] K.S. Jones, S. Walker, and S.E. Robertson, A probabilistic model of information retrieval: development and comparative experiments: Part 2, *Information Processing & Management* **36** (2000), 809-840.
- [8] M. Bendersky and W.B. Croft, Modeling higher-order term dependencies in information retrieval using query hypergraphs, in: *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, 2012, pp. 941-950.
- [9] P. Ogilvie and J. Callan, Combining document representations for known-item search, in: *SIGIR'03*, 2003, pp. 143-150.
- [10] C. Zhai, Statistical language models for information retrieval, *Synthesis Lectures on Human Language Technologies* **1** (2008), 1-141.
- [11] Y. Lv and C. Zhai, Positional language models for information retrieval, in: *Proceedings of the 32nd international ACM SIGIR conference*, 2009, pp. 299-306.
- [12] M. Bendersky, D. Metzler, and W.B. Croft, Effective query formulation with multiple information sources, in: *Proceedings of the fifth ACM international conference on Web search and data mining*, 2012, pp. 443-452.
- [13] D. Metzler and W.B. Croft, A Markov random field model for term dependencies, *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (2005), 472.
- [14] M. Bendersky, D. Metzler, and W.B. Croft, Learning concept importance using a weighted dependence model, in: *Proceedings of the third ACM international conference on Web search and data mining*, 2010, pp. 31-40.
- [15] D. Metzler and W.B. Croft, Latent concept expansion using markov random fields, in: *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, 2007, pp. 311-318.
- [16] F. Diaz and D. Metzler, Improving the estimation of relevance models using large external corpora, in: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2006, pp. 154-161.
- [17] M. Bendersky, D. Metzler, and W.B. Croft, Parameterized concept weighting in verbose queries, in: *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 2011, pp. 605-614.
- [18] E.M. Voorhees and R.M. Tong, Overview of the TREC 2011 Medical Records Track, in: *Proceedings of The 20th Text REtrieval Conference (TREC)*, 2011.
- [19] E. Voorhees and W. Hersh, Overview of the TREC 2012 medical records track, in: *The Twenty-first Text REtrieval Conference Proceedings (TREC)*, 2012.
- [20] D. Zhu, S.T.-I. Wu, J.J. Masanz, B. Carterette, and H. Liu, Using Discharge Summaries to Improve Information Retrieval in Clinical Domain, in: *CLEF (Working Notes)*, 2013.
- [21] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, *Journal of the American Medical Informatics Association* **17** (2010), 507.
- [22] G.K. Savova, J.J. Masanz, P.V. Ogren, J. Zheng, S. Sohn, K.C. Kipper-Schuler, and C.G. Chute, Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications, *Journal of the American Medical Informatics Association : JAMA* **17** (2010), 507-513.
- [23] J.D. Choi and A. McCallum, Transition-based Dependency Parsing with Selectional Branching, in: *ACL (1)*, 2013, pp. 1052-1062.
- [24] D. Metzler and W.B. Croft, A Markov random field model for term dependencies, in: *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, 2005, pp. 472-479.
- [25] B. Croft, D. Metzler, and T. Strohman, *Search Engines: Information Retrieval in Practice*, Addison Wesley, 2009.
- [26] C. Buckley and E.M. Voorhees, Retrieval evaluation with incomplete information, in: *Proceedings of the 27th Annual SIGIR*, 2004, pp. 25-32.
- [27] E. Yilmaz, E. Kanoulas, and J.A. Aslam, A simple and efficient sampling method for estimating AP and NDCG, in: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, 2008, pp. 603-610.
- [28] D. Zhu, S. Wu, B. Carterette, and H. Liu, Using large clinical corpora for query expansion in text-based cohort identification, *Journal of biomedical informatics* **49** (2014), 275-281.
- [29] S. Wu, D. Zhu, W. Hersh, and H. Liu, Clinical information retrieval with split-layer language models, in: *Proceedings of the Association for Computing Machinery SIGIR workshop on health search and discovery (HSD)*, 2013.

Address for correspondence

Stephen Wu, wu.stephen.t@gmail.com, 3181 SW Sam Jackson Park Rd, Portland, OR 97239

A Bayesian Network Model of Head and Neck Squamous Cell Carcinoma Incorporating Gene Expression Profiles

Shuyang Wu^a, Anthony Law, MD, PhD^b, Mark E. Whipple, MD, MS^{a, b}

^aDepartment of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA,

^bDepartment of Otolaryngology - Head and Neck Surgery, University of Washington, Seattle, WA, USA

Abstract

Radiation therapy allows precision targeting of certain groups of lymph nodes and is a treatment for metastatic head and neck squamous cell carcinoma. In current practice, there is approximately 15% probability that physicians inadvertently treat healthy tissue or leave the cancerous lymph nodes untreated. The aim of this work is to improve the accuracy of medical decision-making by extending existing predictive models to capture the probabilities of finding cancerous lymph nodes at each of the six image-based surgical neck level using a patient's genetic profile, primary tumor site and tumor size. Our model was trained with publicly available data extracted from the Cancer Genome Atlas (TCGA) and validated against the TCGA dataset both with and without genetic information. Results show that genetic profile data improves model accuracy. These findings suggest that our predictive model may improve the accuracy of clinical decision-making, especially for patients with more advanced metastasis. However, more data is needed to ensure significance of the proposed effects, as well as to improve accuracy of the overall model.

Keywords: Carcinoma, Squamous Cell; Clinical Decision-Making; Lymph Nodes

Introduction

Head and neck cancers account for approximately 3% of all cancers in the United States [1]. More than 90% of head and neck cancers are squamous cell carcinoma [2]. Head and neck squamous cell carcinomas (HNSCC) have broadly varying survival rates, depending on the primary site, disease stage and the occurrence of metastasis [3]. HNSCC initially metastasizes to the lymph nodes in the neck, following lymph drainage pathways. The regions of the neck containing lymph nodes are classified into six imaging-based surgical neck levels I - VI as shown in Figure 1a [4][5]. If PET or CT scans show evidence that the cancer cells have spread to any of the lymph nodes, radiation therapy can be targeted to treat the area of interest. Radiation therapy is also prescribed if the tumor has reached a certain size, even if there are no detectable signs of lymphatic metastasis. This is because with current technology, small amounts of cancer cells in the lymph nodes do not appear on scans, but there is enough risk in these areas to warrant treatment. The ability to predict locations of lymphatic metastasis is critical for both minimizing the risk of recurrence and minimizing the complications resulting from unnecessary radiation.

Since not all lymph nodes are equally likely to be involved in metastasis, physicians determine which lymph nodes to target based on prior knowledge and personal experience. The decision-making process requires them to estimate many variables including which lymphatic channels the tumor cells

have taken and how far along they have spread within the channels [6]. A study by Crosskerry showed that physician judgment can vary from reality by 15% on average [7]. In other words, there is approximately 15% probability that physicians make either over-conservative decisions and treat the healthy area or leave the cancerous lymph nodes untreated. Several studies have examined the likelihood of certain groups of lymph nodes having cancer based on their prior state. For example, a hidden Markov model was developed based on lymphatic anatomical structure, the primary tumor location and T-stages [6]. A predictive model applying the Bayesian network approach was also established with the same predictors and evaluated to show clinical significance, meaning it could successfully improve accuracy for medical decision-making [8]. With the promising results of both predictive models, we suspected that additional information could be added as a predictor to improve model performance.

Studies have shown that there is a genetic expression profile predictive of nodal metastasis of HNSCC [9][10]. The proposed profile is identified through differential analysis of Affymetrix Human Genome Focus arrays and confirmed by immunohistochemical analysis for transglutaminase-3 and keratin 16 [9][10]. We hypothesized that integrating genetic profile information into our predictive model would improve its performance, leading to more accurate decision aids for clinicians. We were also able to quantify the effects that different genetic profiles play on metastasis by comparing the performance of models with and without genetic information.

We proposed to extend existing predictive models to capture the probabilities of finding cancerous lymph nodes at each of the six image-based surgical neck levels using a patient's genetic profile, primary tumor site and tumor size. Such a model could potentially help physicians to make improved evidence-based decisions while performing targeted treatments such as radiation therapy. The studied effects of a genetic profile on metastasis may enable physicians to adopt more individualized treatments. In addition, future studies can incorporate other risk factors, such as smoking, alcohol and human papillomavirus (HPV) status [11], as predictors in the model via the same method.

Methods

Model construction

The inputs of the model are based on existing research and our hypothesis: tumor origin, the size and local involvement of the primary tumor represented by T-stage, and the patient's tumor-associated genetic profile. The output of the model is the predicted levels of nodal metastasis. Figure 1a shows the surgical neck dissection levels, which are used to delineate regions of cancerous tissue targeted for treatment. Nodal

metastasis tends to follow well-delineated pathways that map to the regions described by the surgical levels.

Based on the physical connectivity of the lymphatic system, cancer cells can only travel to lymph nodes that drain from the primary tumor, and to other lymph nodes through connected drainage channels. We were able to draw from detailed anatomic knowledge on lymphatic connectivity from the Foundational Model of Anatomy (FMA) [12]. We also assumed that despite primary sites or T-stages, once the cancer cells enter the lymphatic system, the metastasis pathway remains consistent and unidirectional, which means the spread of cancer occurs only along described lymphatic channels, and in a direction away from the primary tumor. Clinically, there are rare exceptions due to skip lesions and branching, but because our focus is on studying model performance with additional predictors, we made the decision to exclude these scenarios in the current study. The model outcome is generalized to indicate the highest level of metastatic prognosis. This means for each level below the highest, the estimated probability of that level containing cancer cells is the sum of the probabilities of all levels at or above it.

We developed a Bayesian network model with the structure shown in Figure 1b. In our original model construct, the observed lymphatic level of positive nodal metastasis is conditionally dependent on primary site, T-stage and genetic profile. The weights of the directed edges indicate the probabilistic measures between variables and were obtained by utilizing the available data to learn the network structure in order to validate our assumptions about the predictors. In the learned and final model, the T-stage shows a strong conditional dependence over the primary site. Genetic profile and primary site present a predictive relationship that remains to be investigated.

Data sources

We used two data sources. The first data source was from University of Washington tumor board (UW TB). We obtained records of 383 patients with untreated, non-recurrent squamous cell carcinoma (SCCA) of the head and neck presented to the UW head and neck tumor board over a 3.5-year period. Since this dataset did not contain patient genetic information, we used it to train the baseline model.

The second data source was from the Cancer Genome Atlas (TCGA)[13]. We obtained 528 subjects' clinical data including demographic information, treatment information and disease status, HTSeq-FPKM-UQ (upper quintile of normalized gene expression values), HTSeq-Counts (raw gene expression values) and pathology reports. All data were exported from the TCGA data portal.

TCGA data preparation

Unique identifiers, clinical T-stages, and primary site information were extracted from subjects' clinical data. Because of the limited sample size and the relatively wide range of different primary sites, the number of subjects for each primary site was too small to be statistically significant. Hence, the primary sites were aggregated into larger regional sites as follows: "Tongue", "Floor of Mouth", "Oral Commissure", "Lip", "Alveolar Ridge", "Buccal Mucosa" and "Mandible" were represented as "Oral Cavity"; "Base Of Tongue", "Tonsil", and "Retromolar Trigone" were represented as "Oropharynx"; "Supraglottic larynx" and "Glottic Larynx" were represented as "Larynx". Some primary sites (e.g. "Palate") contained too few samples and were not included. Subjects with unknown T-stage were dropped. Those with stages "T4a" and "T4b" were combined into stage "T4".

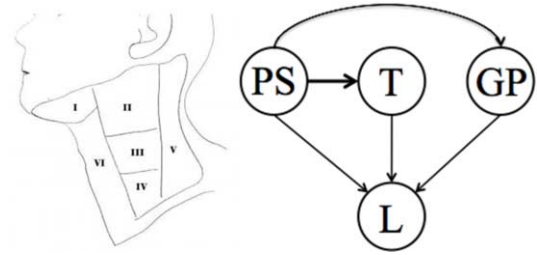


Figure 1a (left). Regions of the neck classified by surgical levels I - VI. Level I, submental and submandibular group; Level II, upper jugular group; Level III, middle jugular group; Level IV, lower jugular group; Level V, posterior triangle group; Level VI, anterior compartment.

Figure 1b (right). Bayesian network model structure. Node PS represents primary site of the tumor; node T represents T-stage/size of the tumor; node GP represents genetic profile of the patient; node I - V represent surgical levels of the neck. The directional arrows indicate conditional relationships between variables. The variables at arrowheads are conditionally dependent on the variables at the tail of the arrows.

Nodal metastatic level information was extracted from pathology reports, linked to corresponding subjects based on the TCGA manifest, and reviewed by two team members for accuracy. Only 349 out of 528 subjects had data on their nodal metastatic level information. Studies performed by Mendez et al. identified differential gene expression profiles of tumor-specific genes in metastatic versus non-metastatic lymph nodes [8][9]; we applied this established gene expression profile to the TCGA dataset. For the purpose of validation, the HTSeq-counts of the subjects in this dataset were also analyzed to identify differentially expressed genes between patients who had lymphatic metastasis and those who did not. The differential analysis was performed using the R "DESeq2" package from Bioconductor (version 3.4). The resulting genes were ranked based on p-value and used to cross-reference with the tumor-specific genes determined in the Mendez et al. studies [8][9]. Of the genes we identified from the TCGA data, 53 genes match genes discovered in the previous Mendez studies. All 53 matched genes were associated with small p-values that fall within the first half of the ranked gene list, thus validating the application of the gene expression profile established from previous studies to the current data set.

Dimension reduction for genetic information

Because the 53 genes are not equally predictive of the level that the lymph nodes were affected, having all of them as individual predictors would mean including those with only a small effect. However, without the biological knowledge on the gene-gene interactions and mechanisms, disregarding any of them using the dimension reduction methods such as PCA and random forest model seemed presumptuous. Therefore, we applied the multivariate logistic regression model published in Dr. Mendez's studies to fit our HTSeq-FPKM-UQ data and the metastatic status [8][9].

The log odds calculated from fitting the logistic regression model were used as propensity scores for each patient to indicate the risk of metastasis associated with his or her genetic profile. The normalized propensity scores of the 349 subjects have the following distribution (figure 2). They were then categorized based on standard deviation to help clinicians to identify patients who are more prone to develop metastasis. Specifically, subjects with propensity scores that are one

standard deviation above the mean are classified as “high risk” ($n = 56$), subjects with propensity scores that are one standard deviation below the mean are classified as “low risk” ($n = 54$), and the rest of the subjects fall under the “regular” risk category ($n = 239$) (supplement figure 2).

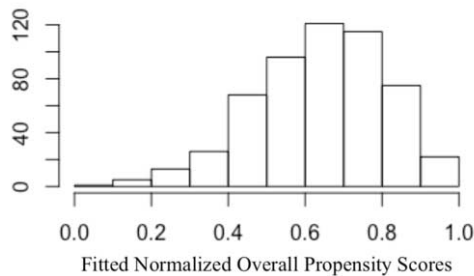


Figure 2 - Histogram of normalized overall propensity scores of the 349 subjects

Parameter Learning/Conditional Probability Table

From the two data sources, TCGA and UW tumor board, we produced four datasets described as follows:

1. TCGA data with genetic expression profile/risk category, sample size = 349, training set = 233.
2. TCGA data without genetic expression profile/risk category, sample size = 349, training set = 233.
3. UW tumor board data without genetic expression profile/risk category, sample size = 383, training set = 255.
4. Combination of TCGA and UW tumor board data without genetic expression profile/risk category, sample size = 482, training set = 321.

Each dataset was trained and tested using the R “bnlearn” package (Version 4.0) with 3-fold cross-validation. When combining the datasets we kept same primary site distributions from the two sources to reduce heterogeneity, data points were randomly selected but reproducible.

Conditional probability tables were produced from fitting Bayesian network models with the training sets above; predictions were made on their corresponding test sets. The model prediction accuracies were measured and compared by calculating the correlation coefficients, cosine similarities and the Area Under the Curve (AUC) values. Because there is no significance test for comparing model accuracies, we decided to use these three measures to ensure the validity of our results.

The correlation coefficient measures the model’s prediction accuracy by correlating the predicted probability of each distinct combination of predictors learned from the training set with the one observed from the test set. Cosine similarity measures prediction accuracy by calculating the cosine of the angle between the predicted level of nodal metastasis and the observed one in the test set. The weighted AUC measures discrimination, that is, the ability of the model to correctly predict subjects’ highest metastatic level. Since our model output has more than two classes, we generated ROC curves for each outcome independently and then averaged them while taking into account the sizes of the classes.

The estimated probabilities of having positive lymph nodes in each level were categorized into “treat,” “maybe treat”, and “do not treat” based on a consensus reached by physicians ($< 7\%$ = do not treat, between 7% and 15% inclusive = maybe treat, $> 15\%$ = treat) as reported in a previous study [14]. We also evaluated the effect of adding genetic information to these decision support tables, and whether it improves clinically

relevant decision-making. The decision support tables were only constructed for subjects with T3 “Oral Cavity” tumors because data on these subjects was most complete and the decision support tables had the least number of missing values.

Results

Table 1 demonstrates the distributions of primary sites, T-stages, and positive nodal metastasis of patients in the training datasets.

Table 1– Datasets Comparison

	TCGA n = 349	UW TB n = 383	Combination n = 280
Oral Cavity	0.70	0.37	0.59
Larynx	0.19	0.20	0.33
Oropharynx	0.10	0.42	0.07
Hypopharynx	0.01	0.00	0.00
T1	0.07	0.18	0.15
T2	0.27	0.34	0.30
T3	0.27	0.19	0.25
T4	0.37	0.28	0.30
No Met	0.51	0.33	0.46
Level I	0.09	0.08	0.08
Level II	0.16	0.28	0.21
Level III	0.13	0.20	0.14
Level IV	0.08	0.08	0.08
Level V	0.03	0.03	0.03

Table 2– Cosine similarities between the datasets

	TCGA vs. UW TB	UW TB vs. Combination	TCGA vs. Com- bination
Cosine similarity	0.86	0.92	0.98

The cosine similarities of different data sets (see Table 2) indicate some heterogeneity between TCGA and UW tumor board data. Specifically, the TCGA dataset contains a larger proportion of subjects with “Oral Cavity” tumors (70% in TCGA; 37% in UW TB), whereas the UW tumor board dataset has a larger proportion (10% in TCGA; 42% in UW TB) of subjects with “Oropharynx” tumor (Table 1). The difference in distribution may have affected our results for clinical decision making comparison as only subjects with “Oral Cavity” cancer were investigated. The comparison of model prediction accuracies learned from the four datasets is shown in Table 3.

Table 3– Model performance and prediction accuracy

Model	genetic info	train size	correlation coefficient	cosine similarity	AUC
TCGA	Y	233	0.80	0.74	0.71
TCGA	N	233	0.70	0.66	0.65
UW TB	N	255	0.63	0.70	0.57
Combination	N	321	0.79	0.73	0.70

The correlation coefficient of the model trained with TCGA data with genetic information and the combination dataset were the highest amongst all the groups. It means the predicted values and the observations of these two models were positively correlated with a coefficient of 0.8.

The model trained with TCGA data with genetic information and the one trained with UW tumor board data both had high cosine similarities, meaning the differences between the predictions and the observations were relatively small.

The weighted AUC result shows that the model trained with TCGA data with genetic information and the model trained with the combination dataset had the highest weighted AUC values, meaning these models resulted in the least amount of classification errors.

Table 4– Conditional Probability Table Comparison for T4 Oral Cavity tumor metastasis

Model	TCGA Whole	TCGA High	TCGA Regular	TCGA Low	UW TB
No Mets	0.43	0.21	0.48	0.43	0.33
Level I	0.16	0.20	0.15	0.15	0.18
Level II	0.21	0.39	0.18	0.21	0.27
Level III	0.08	0.10	0.05	0.15	0.09
Level IV	0.08	0.00	0.10	0.05	0.12
Level V	0.04	0.10	0.05	0.00	0.00

TCGA whole: TCGA data without genetic expression profile/risk category ($n = 233$). TCGA High: TCGA data with high propensity score ($n = 40$). TCGA Regular: TCGA with regular propensity score ($n = 155$). TCGA Low: TCGA with low propensity score ($n = 38$) UW TB: UW Tumor Board ($n = 255$)

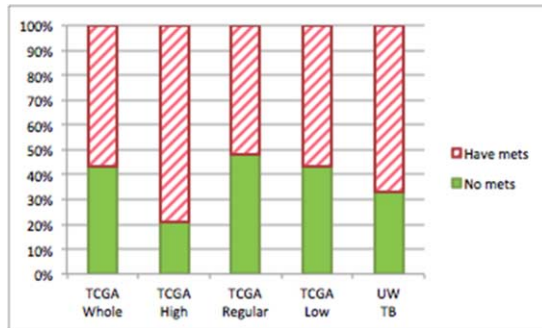


Figure 3 - Comparing the profiles of low, regular, and high-risk patients on metastasis status. Red area represents proportions of subjects in datasets with metastasis; Green area represents proportions of subjects in datasets without metastasis.

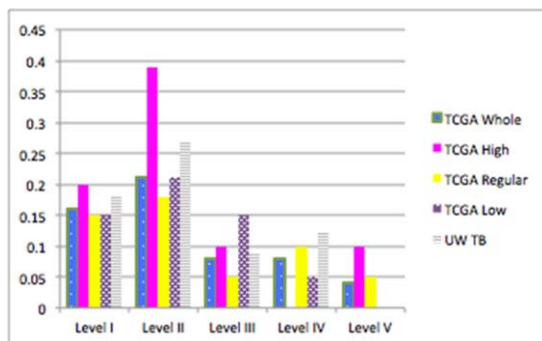


Figure 4 - Comparing the profiles of low, regular, and high-risk patients on metastasis to certain levels. Blue bars represent subjects from TCGA Whole. Pink bar represents TCGA High. Yellow bars represent TCGA Regular. Purple bars represent TCGA Low. Grey bars represent UW TB.

Table 4 shows the comparison of the estimated probabilities of a subject with T4 “Oral Cavity” cancer having metastasis to

specific nodal levels, as predicted by the models trained from each data set.

Probabilities shown in Table 4 and Figure 3 imply that patients with a high propensity score were more likely to develop nodal metastasis regardless of levels (79% of the subjects with “high” risk had metastasis; 57% of the subjects with “low” risk had metastasis). They are also more prone to having level V nodal metastasis (pink bar in Figure 4). However, there was not enough evidence to show significance of these findings because of the relatively small sample sizes, especially for the groups with high and low propensity score.

Table 5– Probability with Treatment Decision Comparison for T4 Oral Cavity tumor metastasis

Model	TCGA Whole	TCGA High	TCGA Regular	TCGA Low	UW TB
No mets	0.43	0.21	0.48	0.43	0.33
Level I	0.57	0.79	0.52	0.57	0.66
Level II	0.41	0.59	0.38	0.41	0.48
Level III	0.20	0.20	0.20	0.20	0.21
Level IV	0.12	0.10	0.15	0.05	0.12
Level V	0.04	0.10	0.05	0.00	0.00

Shading of the cells indicates treatment decisions. Red represents “treat”; yellow represents “maybe treat”; green represents “do not treat”.

Table 5 shows the comparison of the estimated probability of finding positive lymph nodes in each level for subjects with T4 “Oral Cavity” cancer and the treatment decision based on decision thresholds. Comparing to subjects with “regular” risk genetic profiles and subjects from datasets without genetic information, the ones with “high” risk genetic profiles show a change in decision at level V (from “do not treat” to “maybe treat”); subjects with “low” risk genetic profiles show a change in decision at level IV (from “maybe treat” to “do not treat”).

Discussion

All three accuracy measures (correlation, cosine similarity and AUC) showed evidence that having genetic profile as an additional predictor improved our model’s prediction accuracies compared to models trained without genetic information. The improvement seems to be equal to that of having a larger training set when we compared the test results between TCGA ($n = 233$) and Combination ($n = 321$) dataset. This is promising since it’s often more difficult to recruit a larger sample population than to examine all possible predictors in a relatively smaller sample.

Because we used the genetic profile identified in the Mendez et al. [9][10] studies and the associated parameters while preparing the training sets for our model, our results further proved the validity of using the genetic profile to predict metastasis in HNSCC patients. It remains to be investigated whether the genetic profile predictive of metastasis can also be predictive once the cancer cells have spread into the lymphatic system. If there exists a genetic profile that associates with more aggressive cancer cells that lead to higher level of metastasis, it could provide a future research direction.

Due to the sample sizes being too small, we aggregated the primary sites into regions. Because primary sites have differences in metastatic potential even within a region, poorer performance is expected from aggregating into regions than if we had enough data to model individual primary sites. Also, the current model could not reflect the differences in metastatic pathway for each primary site, the results are under the generalized assumption that tumor metastasizes in a linear fashion. For the same reason, even though we were able to

quantify the effects of genetic profile on metastasis, we could not show evidence that such effects are significant and can lead to changes in treatment decisions. However, because the treatment decisions did not change much depending on the dataset used to train the model (see Table 4), it is reasonable to assume that our basic model is consistent in terms of treatment granularity across all data.

The major limitation of this study is the small sample sizes. Because not all pathology reports include the recording of metastatic level, it was difficult for us to gather this data. Data including patient genetic information is even sparser as only certain research studies currently collect patients' genetic information for analysis. Therefore, we needed to use cross-validation rather than using a separate dataset with both genetic profile information and metastatic level information to validate our model. Perhaps future research should include more data for significant results.

Despite the limitations, our study is the first to quantify the effects of genetic profiling on metastasis in HNSCC patients. It is also an indication that genetic information can be used to assist medical decision making in this domain, although the effect size is inconclusive. This calls for increased collection of relevant data to determine significance. Further, we aim to investigate other potential predictors of HNSCC such as smoking, alcohol use, and HPV status to further improve our model performance.

Conclusion

Our predictive model depicted and quantified the relationships between the metastatic level in patients with HNSCC and potential predictors such as tumor primary site, T-stage and metastatic risk propensity derived from patient genetic profiles. Results supported our hypothesis that gene expression has effects on metastasis and including the patient's genetic profile can improve model accuracy. However, there was not enough data to show significance of the effects different genetic profiles have on treatment decision making. All in all, our predictive model can improve scientific knowledge and clinical practice in HNSCC treatment and can be extended to examine more potential risk factors.

Acknowledgements

The authors thank Lucy Wang and Hyunggu Jung for reviewing this paper and providing insightful feedback, and for the support they offered while the study was being conducted.

References

- [1] R.L. Siegel, K.D. Miller, and A. Jemal, Cancer statistics, 2016, *CA Cancer J Clin* **66** (2016), 7-30.
- [2] R.J. Sanderson and J.A.D. Ironside, Squamous cell carcinomas of the head and neck, *Bmj* **325** (2002), 822-827.
- [3] M.J. Ruback, A.L. Galbiatti, L.M. Arantes, G.H. Marucci, A. Russo, M.T. Ruiz-Cintra, L.S. Raposo, J.V. Maniglia, E.C. Pavarino, and E.M. Goloni-Bertollo, Clinical and epidemiological characteristics of patients in the head and neck surgery department of a university hospital, *Sao Paulo Med J* **130** (2012), 307-313.
- [4] P.M. Som, H.D. Curtin, and A.A. Mancuso, Imaging-based nodal classification for evaluation of neck metastatic adenopathy, *AJR Am J Roentgenol* **174** (2000), 837-844.
- [5] K.T. Robbins, J.E. Medina, G.T. Wolfe, P.A. Levine, R.B. Sessions, and C.W. Pruet, Standardizing neck dissection terminology. Official report of the Academy's Committee for Head and Neck Surgery and Oncology, *Arch Otolaryngol Head Neck Surg* **117** (1991), 601-605.
- [6] N. Benson, M. Whipple, and I.J. Kalet, A Markov Model Approach to Predicting Regional Tumor Spread in the Lymphatic System of the Head and Neck, *AMIA Annu Symp Proc* **2006** (2006), 31-35.
- [7] P. Croskerry, From mindless to mindful practice--cognitive bias and clinical decision making, *N Engl J Med* **368** (2013), 2445-2448.
- [8] A. Law, S. Wu, H. Jung, L. Wang, E. Grunblatt, M. Whipple, Prediction of regional metastasis in squamous cell carcinoma of the oral cavity, In preparation.
- [9] E. Mendez, W. Fan, P. Choi, S.N. Agoff, M. Whipple, D.G. Farwell, N.D. Futran, E.A. Weymuller, Jr., L.P. Zhao, and C. Chen, Tumor-specific genetic expression profile of metastatic oral squamous cell carcinoma, *Head Neck* **29** (2007), 803-814.
- [10] E. Mendez, P. Lohavanichbutr, W. Fan, J.R. Houck, T.C. Rue, D.R. Doody, N.D. Futran, M.P. Upton, B. Yueh, L.P. Zhao, S.M. Schwartz, and C. Chen, Can a metastatic gene expression profile outperform tumor size as a predictor of occult lymph node metastasis in oral cancer patients?, *Clin Cancer Res* **17** (2011), 2466-2473.
- [11] R. Sankaranarayanan, E. Masuyer, R. Swaminathan, J. Ferlay, and S. Whelan, Head and neck cancer: a global perspective on epidemiology and prognosis, *Anticancer Res* **18** (1998), 4779-4786.
- [12] C. Rosse, L.G. Shapiro, and J.F. Brinkley, The digital anatomist foundational model: principles for defining and structuring its concept domain, *Proc AMIA Symp* (1998), 820-824.
- [13] Cancer Genome Atlas (TCGA). Genomic Data Commons Data Portal: <https://gdc-portal.nci.nih.gov/>
- [14] I.J. Kalet, M. Whipple, S. Pessah, J. Barker, M.M. Austin-Seymour, and L.G. Shapiro, A rule-based model for local and regional tumor spread, *Proc AMIA Symp* (2002), 360-364.

Address for correspondence

Shuyang Wu, Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA

Email: sw21@uw.edu

Phone: +1-814-441-1004

Applying Risk Models on Patients with Unknown Predictor Values: An Incremental Learning Approach

Enliang Xu^a, Xiang Li^a, Jing Mei^a, Shiwan Zhao^a, Gang Hu^a,
Eryu Xia^a, Haifeng Liu^a, Guotong Xie^a, Meilin Xu^b, Xuejun Li^c

^a IBM Research - China, Beijing, China

^b Pfizer Investment Co. Ltd., Beijing, China

^c Department of Endocrinology and Diabetes, the First Affiliated Hospital, Xiamen University, Xiamen, China

Abstract

In clinical practice, many patients may have unknown or missing values for some predictors, causing that the developed risk models cannot be directly applied on these patients. In this paper, we propose an incremental learning approach to apply a developed risk model on new patients with unknown predictor values, which imputes a patient's unknown values based on his/her k -nearest neighbors (k -NN) from the incremental population. We perform a real world case study by developing a risk prediction model of stroke for patients with Type 2 diabetes mellitus from EHR data, and incrementally applying the risk model on a sequence of new patients. The experimental results show that our risk prediction model of stroke has good prediction performance. And the k -nearest neighbors based incremental learning approach for data imputation can gradually increase the prediction performance when the model is applied on new patients.

Keywords:

Cluster Analysis; Theoretical Models; Risk

Introduction

Risk prediction models are used in clinical decision making to help patients make an informed choice about their treatment. Risk prediction models typically use a number of predictors based on patient characteristics to predict health outcomes. For example, traditional risk models, e.g., CHA2DS2-VASc score [1] and Framingham score [2] for predicting thromboembolism (TE) in Atrial fibrillation (AF), were developed on well-known risk factors that are grounded in previous evidence and experience. Li et al. [3,4] used feature selection methods in machine learning to automatically identify potential risk factors from original features, and achieved more accurate prediction than traditional models. These risk prediction models are developed from well curated registry data.

Advances in information technology has led to the increased availability of large amounts of clinical data, such as the Electronic Health Record (EHR). EHR data contain important information about patients, such as demographics, medical history, medication, lab test results, and billing information. With the prevalence of big data analytics, EHR data is a potentially rich, underutilized source of information that can aid in clinical decision support. The previous work [5] combined expert knowledge with data driven insight for risk factor identification using Electronic Health Records. Works [6,7] used a deep learning approach to extract effective predictors for risk prediction on EHR data.

In clinical practice, the developed risk prediction models cannot be directly applied on new patients, because some of the predictive risk factors are unknown in the new patients. Either the predictor value is missing, or the predictor is not even tested. Data imputation is the process to impute the missing data to yield a more complete dataset. There are various data imputation methods to estimate the missing values [8-10]. These methods commonly use mean, average, or the most frequent values from the whole training data to do imputation.

However, when applying developed risk prediction models on new patients, we do not have training data to estimate the missing values. All we have is the dynamic test data and the patients in test data come sequentially. To address this problem, we propose an incremental learning approach to estimate a patient's unknown predictor values based on k -nearest neighbors (k -NN) [11].

In this paper, we develop a risk prediction model of stroke for patients with Type 2 diabetes mellitus (T2DM) from EHR data. We first preprocess the raw EHR dataset to construct the cohort of interest. Then we use information gain (IG) to identify predictive risk factors, and use logistic regression (LR) to build a risk prediction model based on the selected risk factors. Finally, we propose an approach to apply the developed risk prediction model on new patients with unknown predictor values. We impute missing data using an incremental learning approach based on k -nearest neighbors. The experimental results show that our risk prediction model of stroke has good prediction performance. And the k -nearest neighbors based incremental learning approach for data imputation is effective.

Methods

Figure 1 gives an overview of our method. We first preprocess the EHR data to design the cohort of interest. Then we use feature selection to identify predictive risk factors, and build a risk prediction model based on the selected risk factors. We apply the developed risk prediction model on new patients with unknown predictor values. We impute the missing data using an incremental learning approach based on k -nearest neighbors. In the following, we will describe our method step by step in more detail.

EHR Data

EHR data contain important information about patients, such as demographics, medical history, medication, labtest results, and billing information. We conduct experiments on a real world EHR dataset collected from August 1 2011 to March 31 2016.

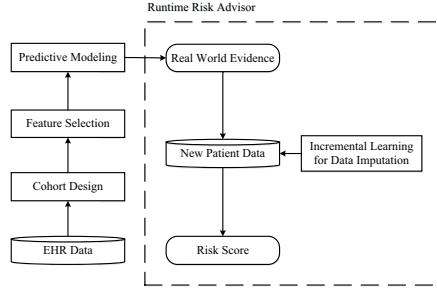


Figure 1– Method Overview.

The multi-relational EHR data involves four medical entities (patient, diagnosis, medicine and labtest), which constitutes major information to describe patients' condition. The data is actually a collection of medical events where each event can be described as a patient p required to take a medicine m for a diagnosis d .

We first preprocess the EHR data. In the raw EHR data, a large amount of features have non-standardized and dirty values. Many features are missing due to unknown values or errors in data collection. We perform data cleansing and missing data normalization before predictive analysis. The raw EHR data is heterogeneous and includes different types of data items: binary type (e.g., medication), nominal type (e.g., gender) and numeric type (e.g., labtest). For different data types, we standardize data formats, correct input errors, and discard the values that are not structured as the target types. As the dataset is collected from different hospitals, a numeric feature may have different units, we unify these features to the same unit. Finally, we also discard the numeric values that are out-of-range.

After data cleansing and missing data normalization, the patient entry consists of 7 demographic features, i.e. "gender", "age", "blood", "blood_rh", "education", "occupation", "marriage". The diagnosis entry consists of 270 ICD 10 codes. The medicine entry consists of 87 drug names. The labtest entry consists of 124 labtests. For diagnosis, medicine, and labtest, we remove features with a frequency less than 1%. There are 489 features in the training and test set including patient id.

Cohort Design

We define the index event as the stroke onset of a patient. The cohort is T2DM without stroke observed before the index event. The outcome is stroke. The ICD 10 codes for stroke are [I60, I69]. The number of cases and controls are 17,852 and 93,909, respectively.

We set the observation window to 2 years before the index event and the prediction window to 1 year after the index event. Figure 2 shows the cohort design.

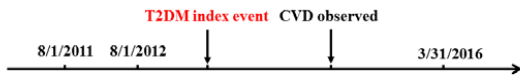


Figure 2– Cohort Design.

To preserve the distribution of cases and controls in the training and test set, we randomly select 70% of the cases and controls as a training set, and use the remaining 30% of the cases and controls as a test set. There are 78,232 patients in the training set and 33,529 patients in the test set.

Since the dataset is derived from real world EHR data, the labtest entry is very sparse. To build a reasonable risk

prediction model, we normalize the missing data in the labtest entry with the average value for each feature.

Feature Selection

Feature selection is the process of selecting a subset of relevant features for use in model construction. We use the filter strategy to identify predictive risk factors. The filter strategy calculates a score to represent the relevance of each feature against the outcome, and then ranks and filters the features based on the score. In this study, we use information gain (IG) as the relevance score, which is the change in information entropy when a feature is given.

Predictive Modeling

We use logistic regression (LR) to build a risk prediction model based on the selected risk factors. LR is a generalized linear model with a logit link function and a binomial distribution, which is widely used in both medical statistics and machine learning, because of its good prediction ability and interpretability.

The standard metric AUC is used to evaluate the performance of the risk model based on information gain (IG) selected features. In addition, we also compare AUPR, the area under the precision recall curve, with a baseline model. In many cases of risk prediction, the number of positive samples are much smaller than that of negative ones. For these highly imbalanced datasets, AUC may provide an overly optimistic view of performance, while AUPR can provide a more informative assessment [12]. We also report the accuracy of our risk prediction model.

Incremental Learning for Data Imputation

In clinical practice, developed risk prediction models cannot be directly applied on new patients, due to unknown predictor values. Data imputation is the process to remedy missing data, which is usually necessary for building a reasonable prediction model. In this study, we propose an incremental learning approach to impute the missing data based on k -nearest neighbors (k -NN). Specially, when a new patient comes, we first find his/her k -nearest neighbors using non-missing features from existing patients.

Here we use the commonly used Euclidean distance as the distance metric. The Euclidean distance between patient x and y is defined as follows.

$$d(x, y) = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (1)$$

where x_i and y_i is the i^{th} feature of patient x and y respectively.

As the labtest data is very sparse, we normalize the distance by the number of non-missing features that both patient x and y has.

$$d(x, y) = \frac{\sqrt{\sum_{i=1}^k (x_i - y_i)^2}}{k} \quad (2)$$

Given a patient, we first find his/her k -nearest neighbors using the normalized Euclidean distance. Then we use the average value of the non-missing features of the k -nearest neighbors to impute the missing feature of the patient. In this way, we can reduce the bias introduced by the imputation.

For example in Table 1, suppose patient P_1 , P_2 and P_3 are the 3-nearest neighbors of patient P_4 . The underlined feature value in the table indicates that the feature is missing and its value is imputed. When imputing the missing feature a of patient P_4 , we do not use the feature a value of patient P_2 , as it is imputed. Instead, we use the average value of feature a of patient P_1 and P_3 to impute that of patient P_4 , which is 4.

Table 1– *K-nearest Neighbor based Data Imputation Example.*

Feature	P ₁	P ₂	P ₃	P ₄
a	5	<u>4</u>	3	<u>4</u>
b	<u>2</u>	5	3	6
c	6	3	<u>5</u>	2

When the first patient comes or a patient has no k -nearest neighbors, we use prior knowledge, e.g. a feature's normal value, to impute the missing feature of the patient.

Using the average feature value of a patient's k -nearest neighbors to impute a missing feature does not consider the importance of different k -nearest neighbors. To address this problem, we can use the weighted average value of k -nearest neighbors to do the imputation as shown below.

$$x^a = \frac{\sum_{i=1}^k S_{xy_i} y_i^a}{k} \quad (3)$$

where x^a is the feature a value of patient x , S_{xy_i} is the similarity between patient x and y_i , and y_i^a is the feature a value of patient y_i .

We can also combine the weighted average value of a patient's k -nearest neighbors and prior knowledge to do the imputation by introducing a parameter π . Parameter π controls the importance of the weighted average feature value of k -nearest neighbors and the feature's normal value.

$$x^a = \pi * \frac{\sum_{i=1}^k S_{xy_i} y_i^a}{k} + (1 - \pi) * K(a) \quad (4)$$

where $K(a)$ is the normal value of feature a from prior knowledge.

Results

We perform a case study using our method on the real world EHR data. We use logistic regression (LR) to build a risk prediction model from a training set based on information gain selected risk factors. The number of predictors is set to 30. We have tried different numbers of predictors and found that the models have similar performance. Three standard metrics, AUC, AUPR and Accuracy (ACC), are used to evaluate the prediction performance of our model. In the experiment, we set the threshold to 0.5.

We first compare the prediction performance of our model with another two models. One model is built using all the features in the dataset. Another model is built using only the age feature.

Table 2 shows the AUC, AUPR and accuracy (ACC) results of different models. We can see that our model has better prediction performance than the other two models, as the AUC, AUPR and accuracy of our model are all higher than that of the other two models. The AUC and accuracy of our model is slightly better than that of the model using all features, indicating that the features selected by information gain (IG) are more relevant against the outcome, i.e. stroke. Top predictors selected by information gain with corresponding scores are shown in Table 3.

Table 2– *AUC, AUPR and ACC Results of Different Risk Models based on Different Feature Sets.*

Model	AUC	AUPR	ACC
Model using age feature	0.6916	0.2545	0.8397
Model using all features	0.7708	0.3872	0.84
Model using IG features	0.7859	0.464	0.8548

Table 4 shows the evaluation results (AUC, AUPR and Accuracy) of our model using incremental learning for data imputation by varying the number of patients from 10 to 1000. Next, we compare k -nearest neighbor based data imputation (EHR-based model_knn) with an existing data imputation method, i.e. using the average value to handle missing data (EHR-based model_avg).

We first plot the AUC value for every 100 patients as shown in Figure 3. We can see that for the first 100 patients, the average value based data imputation has better AUC than the k -nearest neighbor based method. However, after 100 patients, k -nearest neighbor based data imputation performs better.

For k -nearest neighbor based data imputation, there is a drastic increase in AUC value from 10 to 200 patients. After 200 patients, the AUC value tends to be stable. The turning point at 200 patients indicates that as the number of incoming patients increases, the k -nearest neighbor based missing data imputation will be more accurate before 200 patients. However, after 200 patients, more incoming patients do not help to improve the AUC value significantly.

We also find that as the number of incoming patients increases, the runtime of imputing the missing feature values of a patient based on k -nearest neighbors also increases. Therefore, when the AUC value gets stable, the patient set is enough to find k -nearest neighbors for new incoming patients to do imputation, saving runtime and resources.

Table 3– *Top Predictors Selected by Information Gain.*

Feature Name	Feature Score
Cholinesterase	0.4392
MCHC	0.4381
Lipoprotein(a)	0.4368
Na	0.4368
Bacteria	0.4363
Lactate dehydrogenase	0.4265
Prealbumin	0.4227
AST	0.4227
Total protein	0.415
Uric acid	0.4124
Age	0.0536

We further zoom in to investigate the AUC value from 10 to 200 patients. As can be seen in Figure 4, the k -nearest neighbor based method has better or comparable AUC than the average value based method. When the number of incoming patients is less than or equal to 200, more patients will help improve the AUC value using incremental learning for data imputation based on k -nearest neighbors.

Table 4– *AUC, AUPR and ACC Results by Varying the Number of Patients from 10 to 1000.*

#PATIENTS	AUC	AUPR	ACC
10	0.5714	0.2381	0.7778
20	0.6833	0.2909	0.7895
30	0.6364	0.3235	0.7586
40	0.6855	0.3542	0.7949
50	0.6962	0.4762	0.7755
60	0.7397	0.5725	0.7797
70	0.6757	0.4956	0.7826
80	0.684	0.4826	0.7848

90	0.6877	0.5024	0.7865
100	0.6778	0.4206	0.8081
110	0.665	0.3733	0.8165
120	0.6617	0.3399	0.8235
130	0.6589	0.3061	0.8295
140	0.6683	0.3032	0.8345
150	0.6915	0.3605	0.8389
160	0.7041	0.3923	0.8491
170	0.7133	0.4479	0.8521
180	0.7169	0.4399	0.8603
190	0.7255	0.4381	0.8624
200	0.741	0.4659	0.8643
300	0.7383	0.4402	0.8595
400	0.7511	0.4895	0.8546
500	0.7447	0.4417	0.8577
600	0.7397	0.4379	0.8564
700	0.7539	0.462	0.8655
800	0.7432	0.4499	0.8673
900	0.7392	0.4411	0.8654
1000	0.7369	0.4341	0.8659

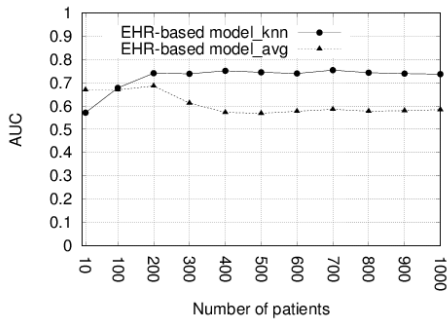


Figure 3– AUC for Every 100 Patients by Varying the Number of Patients from 10 to 1000.

As in our dataset, the number of positive samples are much smaller than that of negative ones, AUC may provide an overly optimistic view of performance, while AUPR can provide a more informative assessment. So we report the AUPR for every 100 patients in Figure 5.

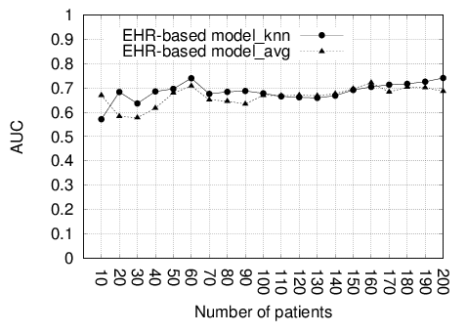


Figure 4– AUC for Every 10 Patients by Varying the Number of Patients from 10 to 200.

We compare the AUPR of k -nearest neighbor based data imputation, average value based data imputation with a baseline, which is the average precision of randomly predicting the risk. We can see that the AUPR of the k -nearest neighbor based method increases from 10 to 200 patients, and tends to be stable after 200 patients. The k -nearest neighbor based method outperforms the average value based method after 100 patients. The AUPR of average value based data imputation decreases as the number of incoming patients increases. Both methods have better AUPR than the baseline.

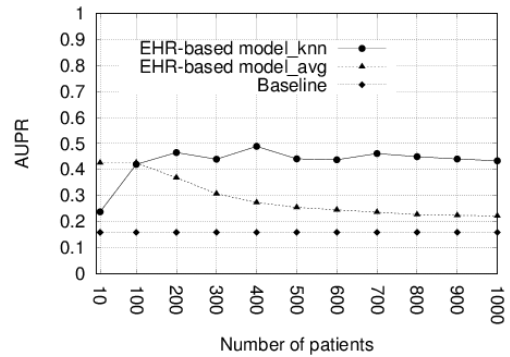


Figure 5– AUPR for Every 100 Patients by Varying the Number of Patients from 10 to 1000.

As with the AUC, we also report the AUPR for every 10 patients as shown in Figure 6. We can see that the AUPR of both k -nearest neighbor based data imputation and average value based data imputation are not stable for the first 200 patients.

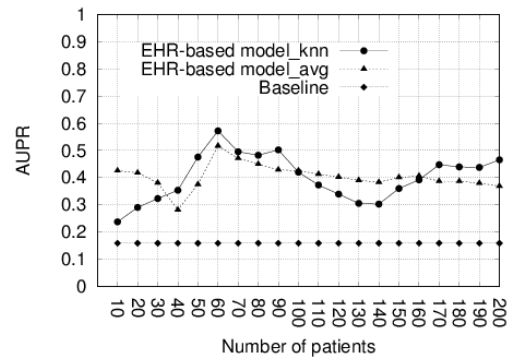


Figure 6– AUPR for Every 10 Patients by Varying the Number of Patients from 10 to 200.

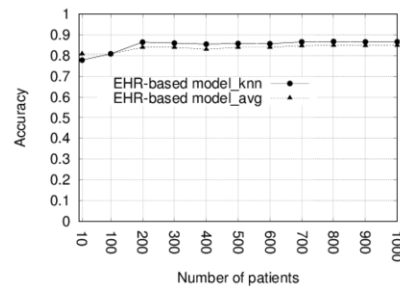


Figure 7– Accuracy for Every 100 Patients by Varying the Number of Patients from 10 to 1000.

We also evaluate the prediction performance by accuracy. Figure 7 shows the accuracy for every 100 patients. We can see that there is a slight increase in accuracy from 10 to 200 patients for both methods. After 200 patients, the accuracy does not change much. However, the k -nearest neighbor based method has better accuracy than the average value based method.

Figure 8 shows the accuracy for every 10 patients. We can see that the k -nearest neighbor based data imputation and average value based data imputation have comparable accuracy. And the accuracy value of both methods increases as the number of patients increases.

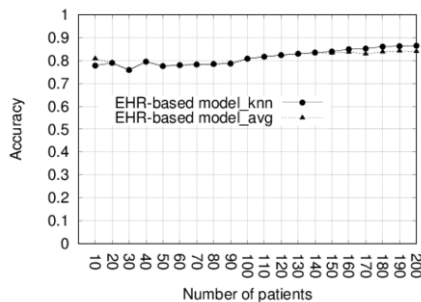


Figure 8—Accuracy for Every 10 Patients by Varying the Number of Patients from 10 to 200.

Discussion

We propose an incremental learning approach to apply a developed risk model on new patients with unknown predictor values, which imputes a patient's unknown values based on his/her k -nearest neighbors (k -NN) from the incremental population. As a new patient comes, we find his/her k -nearest neighbors from all the existing patients to estimate the missing values. There is a trade off in this approach. With more existing patients, the k -nearest neighbors we find will be more accurate, leading to a better estimate of the missing values. However, it is also very time consuming. In order to balance accuracy and efficiency, we can use a sliding window to keep the most recent patients to find k -nearest neighbors. k -nearest neighbor based missing data imputation also requires good data quality. If the patient data is very sparse, we cannot even find the k -nearest neighbors for a new patient.

We perform a real world case study by developing a risk prediction model of stroke for patients with T2DM from EHR data. The sparse, heterogeneous, and noisy data stored in the EHR present a multitude of technical challenges for analysis. EHR data contain many heterogeneous medical events, such as diagnosis, procedures, medications and labtests, making the dimensionality of the EHR very high. EHR data is also very noisy due to non-standardized physician practices, missing values, etc. As a result, some features selected from the EHR data may be uninterpretable, not classic, or even irrelevant for CVD. So before feature selection, it would be better to perform pre-selection to remove unreasonable, duplicate features. Non-discriminative features, in which most of the instances have identical values, should also be removed. Domain knowledge is also important; with the help of clinicians, we can select features of interest, and discard others.

When developing the risk model, we use ICD 10 codes [I60, I69] for the outcome stroke. However, ICD coding is incorrect between 23% and 40% of the time. Although T2DM is a risk factor for stroke, the time course and severity of illness are necessary to know whether stroke is truly related to CVD.

Our incremental learning approach for missing data imputation gives a direction to applying developed risk prediction models on new patients with unknown predictors. Here we use k -nearest neighbors to handle missing data. Other data imputation methods can also be used.

Conclusion

In this paper, we proposed an incremental learning approach to apply a developed risk model on new patients with unknown predictor values, which imputes a patient's unknown values using his/her k -nearest neighbors (k -NN) from the incremental population. We performed a real world case study by developing a risk prediction model of stroke for patients with T2DM from EHR data, and incrementally applied the risk model on a sequence of new patients. The experimental results show that our risk prediction model of stroke has good performance. The k -nearest neighbor based incremental learning approach for data imputation can gradually increase the performance when the model is applied on new patients.

References

- [1] GY Lip, R Nieuwlaat, R Pisters, D A Lane and H J G M Crijns. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest*, **137**(2010) 263-72.
- [2] T J Wang, J M Massaro, D Levy D, R S Vasan, P A Wolf, R B D'Agostino, M G Larson, W B Kannel and E J Benjamin. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: the Framingham heart study. *JAMA*, **290** (2003) 1049-56.
- [3] X Li, H Liu, X Du, P Zhang, G Hu, G Xie, S Guo, M Xu and X Xie. Integrated machine learning approaches for predicting ischemic stroke and thromboembolism in atrial fibrillation. *AMIA Annu Symp Proc*: (2016) 799-807.
- [4] X Li, H Liu, X Du, G Hu, G Xie and P Zhang. Using frequent item set mining and feature selection methods to identify interacted risk factors - the atrial fibrillation case study. *Studies in Health Technology and Informatics*, **228** (2016) 562-566.
- [5] J Sun, J Hu, D Luo, M Markatou, F Wang, S Edabollahi, S E Steinhilb, Z Daar and W F Stewart. Combining knowledge and data driven insights for identifying risk factors using electronic health records. *AMIA Annu Symp Proc*. 2012:901-10.
- [6] E Choi, A Schuetz, W F Stewart and J Sun. Using recurrent neural network models for early detection of heart failure onset. *J Am Med Inform Assoc* **24** (2017), 361-370.
- [7] Y Cheng, F Wang, P Zhang and J Hu. Risk prediction with electronic health records: a deep learning approach. *SIAM International Conference on Data Mining*, 2016.
- [8] B J Wells, K M Chagin, A S Nowacki and M W Kattan. Strategies for handling missing data in electronic health record derived data. *EGEMS*, **1** (2013) 1035.
- [9] J M Engels and P Diehr. Imputation of missing longitudinal data: a comparison of methods. *J Clin Epidemiol*, **56** (2003) 968-976.
- [10] P García-Laencina, J Sancho-Gómez and A Figueiras-Vidal. Pattern classification with missing data: a review. *Neural Comput Appl*, **9** (2009) 1-12.
- [11] J Read, A Bifet, B Pfahringer and G Holmes. Batch-incremental versus instance-incremental learning in dynamic and evolving data. 11th Adv. Intell. Data Anal., 2012, pp. 313-323.
- [12] H B He and E A Garcia. Learning from imbalanced data. *IEEE Transactions of Knowledge and Data Engineering*, **21** (2009) 1263-84.

Address for correspondence

Enliang Xu, Ph.D, IBM Research – China

Building 19, Zhongguancun Software Park, Haidian District, Beijing, China, 100193

Email: bjaxlxu@cn.ibm.com

Semantic Relations in Compound Nouns: Perspectives from Inter-Annotator Agreement

Prabha Yadav^a, Elisabetta Jezek^b, Pierrette Bouillon^c, Tiffany J. Callahan^a,
Michael Bada^a, Lawrence E. Hunter^a, and K. Bretonnel Cohen^a

^a Computational Bioscience Program, University of Colorado School of Medicine, Aurora, Colorado 80045, USA

^b Department of Humanities, University of Pavia, Italy

^c Faculté de Traduction et d'Interprétation, Université de Genève, Switzerland

Abstract

Semantic relations have been studied for decades without yet reaching consensus on the set of these relations. However, biomedical language processing and ontologies rely on these relations, so it is important to be able to evaluate their suitability. In this paper we examine the role of inter-annotator agreement in choosing between competing proposals regarding the set of such relations. The experiments consisted of labeling the semantic relations between two elements of noun-noun compounds (e.g. cell migration). Two judges annotated a dataset of terms from the biomedical domain using two competing sets of relations and analyzed the inter-annotator agreement. With no training and little documentation, agreement on this task was fairly high and disagreements were consistent. The results support the utility of the relation-based approach to semantic representation.

Keywords:

Natural Language Processing; Evaluation Studies as Topic

Introduction

Linguists have tried to discover the basic building blocks of semantic relations between nouns for decades, but there is still little consensus about what the set of those building blocks might be. This is an important problem for biomedical natural language processing and biomedical ontologies, because those building blocks are at the heart of our information extraction tasks and the structure of our ontologies.

Compound nouns are crucial to practical tasks like knowledge representation and to theoretical problems like understanding compositionality in semantics [1; 2]. However, one of the most difficult problems in semantic representation and in language processing is the nature of the relations between the two parts of a compound noun [3-6] (see examples in Table 1). Compound nouns are formed by a sequence of two or more nouns [7]. In writing, they may appear as two tokens (*knockout mouse*), a hyphenated word (*nucleotide-excision*), or a single token (database) [7]. They are about twice as common in written English as they are in spoken English (248/million words in newswire text versus 123/million words in conversation)[7]. They are quite common in scientific writing. Linh (2010) [8] reviews a number of studies of the incidence of compound nouns in technical texts, reporting that one study found that 27% of words in scientific abstracts were in compound nouns; another study found that 11.86% of anaphors are compound nouns; and another found that

15.37% of a technical corpus was made up of compound nouns.

The study of compound nouns dates back to Pāṇini and Kātyāyana and Patañjali [9], but an enormous amount of work remains to be done, particularly on the semantics of the relations between the nouns in a compound, and they remain the topic of a considerable amount of research in both linguistics and natural language processing [3]. Various authors have attempted to describe the relation between the elements of compound nouns from a theoretical perspective [10]. Likewise, a number of studies in language processing have shown the difficulty of classifying the relations in these compounds automatically [11-15]. All of these studies are based on specific representations of the relations that can hold between the nouns in a compound. This raises a question: are those relations valid? One way to answer that question is by measuring whether humans can reliably label the relation that holds in any specific compound. If they cannot, then we must question the validity of the representation itself, and we must consider the possibility that any principled investigation of the relations in compounds, whether from a theoretical or a practical perspective, is impossible (see, for example, the logical positivist perspective and how inter-annotator agreement responds to the problems of “observing” semantics [16]. On the other hand, if they can, then it might be possible to train computers to do the same task, which could enable considerable advances in natural language processing.

We can address the reliability of labeling through examining inter-annotator agreement when two or more analysts label the relations in a sample of compound nouns. However, we are not aware of any studies that have looked at inter-annotator agreement in compound nouns. Identifying relations in compound nouns, whether done by humans or by computers, is a non-trivial task because there is an enormous amount of ambiguity in the correspondence between semantics and syntactic structure. For example, Table 1 gives a number of examples using the biomedical term *forceps*. We note that *forceps* can exist in at least five relations with another noun in a compound—that is to say, the same noun-noun syntactic structure can correspond to at least five relations between the first noun and *forceps*.

Table 1- Identical syntactic structures can reflect a wide variety of semantic relations

Relation	Example
Used_on	<i>Bone forceps</i>
Instrument_for	<i>Epilation forceps</i>
Shape_of	<i>Mosquito forceps</i>
Operated_by	<i>Thumb forceps</i>
Named for	<i>Kelly forceps</i>

Consider the term *chondrocyte development*. The relation between the two nouns is an activity/physical process—development—that is undergone by chondrocytes. *Motor activity* and *thrombin activity* have the same syntactic structure as *chondrocyte development* and as each other, but the relationships between the nouns in all three of them are different: an activity that is undergone in the first, the result of the second, and the action of the third.

This article investigates the ability of humans to reliably label semantic relations between the elements of noun compounds in the face of this semantic ambiguity in identical syntactic structures. The motivation for this is that inter-annotator agreement on labelling the semantic relations in compound nouns is a useful indicator of the validity of the proposed set of relations and can be used in choosing between two competing theories. The experiment was done using two sets of relations with two different contexts of theoretical status and computational applications—Generative Lexicon theory on the one hand, and a model of the domain on the other—holding constant the data set and the annotators. Good inter-annotator agreement for a given set of relations would lend some credibility to that set; bad inter-annotator agreement would detract from its credibility.

Materials

Generative Lexicon

The first set of relations is the Generative Lexicon relations described in Bouillon et al., 2012 [17]. We will refer to this set as GL relations in the following, although the set in Bouillon et al. includes extensions with respect to the original GL set [18], for example the tag *argumental*. GL theory is an attempt to explain how compositionality contributes to lexical semantics. Bouillon et al., 2012 [17] posit two basic elements of lexical semantic representations: Qualia and/or Argumental. Qualia relations involve predicates and their arguments, as well; we will re-visit this issue in the discussion. They identified four basic Qualia relations: Formal, Constitutive, Telic, and Agentive. We used the set of Generative Lexicon relations described in Bouillon et al., 2012 [17]. These relations are meant to be general and elementary, embodying a hypothesis about the fundamental building blocks of semantic representations.

Rosario and Hearst

Rosario and Hearst (2001) [4] identified 38 relations broadly inspired by linguistic theory, but the motivation for the relations is less language-theoretic and more application-oriented. Specifically, it is intended to represent the semantics and knowledge structures of biomedical literature. Where GL theory is meant to be cross-linguistically valid, the set of relations proposed by Rosario and Hearst is meant to be domain-specific—thus, it does not attempt to be valid even for all of the English language, but just for English-language scientific literature. In comparison to many other proposed sets of semantic relations, including that of Generative

Lexicon theory (Table 2), the Rosario and Hearst relations do not posit a set of semantic primitives. Rather, they embody a knowledge representation schema that is specifically tailored to biomedical science without making any claims about what relations might exist in other domains.

Table 2- The characteristics of the two sets of relations in terms of their size, goals, and generality/domain specificity

Relation set	Size	Goal	General	Domain-specific
Generative Lexicon	15	Theoretical	X	
Rosario and Hearst	38	Application-oriented		X

The sample of terms for annotation was drawn from the Gene Ontology (GO). The GO is a good test case for this study because its content is clearly relevant to the biomedical domain and because it contains a large number of noun compounds. The GO project (<http://geneontology.org/>) was founded in 1998. The GO is an open-access, community-based effort to create a unified representation of three biological concept domains (biological processes, cellular components, and molecular functions) that are used to describe the activity of genes and gene products in a species-independent manner. Each of the 41,775 GO concepts includes a term name [19; 20].

Methods

To select the sample, all of the terms in the GO were tagged with their part of speech using the CLEAR suite of language processing tools [21; 22]. Tagging errors are noted in the data set. All terms with exactly two words, such that both words are nouns—that is, compound nouns—were pulled from the full set. We then selected a random sample of 101 words from the compound nouns (intending 100, with an extra in case of tagging errors).

Annotators

One annotator has a bachelor's degree in biology. The other annotator is a cardiovascular technologist with a doctorate in linguistics. They were instructed to base their annotations of the relation in a term on the definition of that term on the GO web site.

Evaluation

The inter-annotator agreement was measured using F-measure and Cohen's Kappa. In the calculation of kappa, P(e) (expected chance agreement) was calculated from the marginals of the confusion matrix. This is a conservatively high estimate of the P(e), and that should be kept in mind when interpreting the results.

Results

The Generative Lexicon relations from Bouillon et al.

The annotators used eight of the 15 possible relations to annotate the 101 GO terms (see Table 3). The most commonly used relation by annotator 1 was *argument* followed with *played_by*. The most commonly used relation by annotator 2 was *played_by* followed with *used_for*. Annotator 1 thought there were no proper relations available for four terms: *larval development*, *predatory behavior*,

lymphocyte energy, and *lymphocyte homeostasis*. Table 4 shows the results in terms of true positives (1 for each match between the two annotators), false positives (1 for each mismatch between the two annotators), and false negatives (also 1 for each mismatch between the two annotators), and the corresponding measures of inter-annotator agreement. Cohen's Kappa value was 0.47 and the inter-annotator agreement, calculated as F-measure, was 0.58. The Cohen's Kappa value indicates a fair/good level of reliability according to the Green scale (1997). The annotators agreed that the Telic relation was the most frequent relation, followed by the Argumental relation. Annotator 1 thought 54.45% of the terms were Telic and annotator 2 thought 70.29% of the terms were Telic (Table 3). 36.63% of the terms were annotated as Argumental by annotator 1 and 19.80% of the terms were annotated as Argumental by annotator 2.

As defined by Cohen (1960) [23], unweighted kappa was calculated using the following equation (Eq. 1):

$$k = \frac{p_o - p_e}{1 - p_e} \quad (1)$$

Solutions were verified using the irr R package [24]

Table 3- Distribution of relations using GL relations

Relation	Annotator 1 (%)	Annotator 2 (%)	Examples
Formal [is_a]	0	0.99	<i>predatory behavior</i>
Constitutive [made_of]	0.99	0.99	<i>dynactin motor</i>
Constitutive [member_of]	0.99	0	<i>kinin cascade</i>
Telic [used_for]	0	20.79	<i>chondrocyte differentiation,</i>
Telic [aims_at]	21.78	18.81	<i>translation reinitiation,</i>
Telic [played_by]	32.67	30.69	<i>GTPase activity,</i>
Agentive [caused_by]	2.97	7.92	<i>chondrocyte hypertrophy, heart wedging</i>
Argumental [Argument]	36.63	19.80	<i>protease binding, p53 binding</i>
Un-annotated	3.96	0	<i>lymphocyte homeostasis</i>

Table 4 Overall inter-annotator agreement using GL relations

TP	FP	FN	P	R	F	κ
59	42	42	0.58	0.58	0.58	0.47

Table 5- Bouillon et al. GL categories, relations, descriptions, Relation Ontology equivalent(s), and examples

GL Category	Relations	Descriptions	RO equivalent	Examples
Formal	is_a	N2 is a kind of N1	is_a/ subclass of	<i>predatory behavior</i>
	shape_of	N1 has the shape of N2		
	holds	N1 holds N2	Contains	
Constitutive	made_of	N1 is made of N2	has_proper_part	<i>dynactin motor</i>
	part_of	N1 is a part of N2	proper_part_of	
	located_in	N1 is spatially/temporally located in N2	located_in	
	member_of	N1 is member of N2	member_of	<i>kinin cascade</i>
	has members	N1 has N2 as members	has_member	
Telic	predicate	N1 has the purpose of (Predicate) N2	has_function	
	used_for	N1 is used for the activity N2		<i>chondrocyte differentiation</i>
	aims_at	N1 has N2 as result/end goal		<i>keratinocyte development</i>
	played_by	N1 denotes the function which is N2.	has_function	<i>cholinesterase activity</i>
Agentive	caused_by	N1 is created/brought into existence/caused by N2		<i>heart wedging, chondrocyte hypertrophy</i>
	derived_from	N1 is derived/extracted from N2	derives_from, transformation_of	
Argumental	argument	N2 is an argument of N1		<i>p53 binding</i>

Table 6¹ shows the confusion matrix for these relations. Note that disagreements between the two annotators are largely systematic. For example, Annotator 1 and Annotator 2 used the *aims_at* relation a similar amount of times (22 and 19), agreeing in the case of 10 noun compounds (Annotator 1 10/22 (45.45%); Annotator 2 10/19 (52.63%)). Of the disagreements, 8/22 times that Annotator 1 labelled instances of the *aims_at* relation, Annotator 2 labelled the *used_for*

relation; 7/19 times that Annotator 2 labelled the *aims_at* relation, Annotator 1 annotated the *argument* relation. Thus, refining the guidelines such that it is clearer when to use *aims_at* versus *argument* and *used_for* would have a large effect on the inter-annotator agreement for all three of these relation types.

The Relation Ontology relations corresponding to the Generative Lexicon relations are shown in Table 5. We observed that there were no corresponding ontology relations for most of the Generative Lexicon relations. This is consistent with the suggestion that the Relation Ontology is missing content that is fundamental to representing the biomedical domain.

¹ Table 6 is on the GitHub Repository: <https://github.com/KevinBretonnelCohen/SemanticRelationsCompoundNouns.git/>

Table 7- Rosario and Hearst relations and examples. There are no Relation Ontology equivalents for these relations.

Name	Examples	Name	Examples
Wrong parse	exhibit asthma, ten drugs	Time of (1-2)	morning headache, hour headache
Subtype	headaches migraine, fungus candida	Measure of	relief rate, asthma mortality, asthma morbidity
Activity/Physical process	bile delivery, virus reproduction, bile drainage, headache activity	Person/center who treats	headache specialist, headache center, diseases physicians, asthma nurse
Ending/reduction	migraine relief, headache resolution	Instrument (1-2)	aciclovir therapy, chloroquine treatment
Defect in Loc.	lung abscess, artery aneurysm	Instrument (2-1)	vaccine antigen, biopsy needle
Change	papilloma growth	Instrument (1)	heroin use, internet use, drug utilization
Produces (on a genetic level)	polyomavirus genome, actin mRNA, CMV DNA, protein gene	Object	bowel transplantation, kidney transplant, drug delivery
Cause (1-2)	asthma hospitalizations, AIDS death	Misuse	drug abuse, acetaminophen overdose
Cause (2-1)	flu virus, diarrhea virus	Subject	headache presentation, glucose metabolism
Characteristic	receptor hypersensitivity, cell immunity	Purpose	headache drugs, HIV medications
Physical property	blood pressure, artery diameter	Topic	time visualization, headache questionnaire
Defect	hormone deficiency, CSF fistulas	Location	brain artery, tract calculi, liver cell
Physical Make Up	blood plasma, bile vomit	Modal	emergency surgery, trauma method
Person afflicted	AIDS patient, BMT children	Material	formaldehyde vapor, aloe gel, gelatin powder
Demographic attributes	childhood migraine, infant colic, women migraineur	Frequency/time of (2-1)	headache interval, attack frequency, football season
Bind	receptor ligand, carbohydrate ligand	Activator (1-2)	acetylcholine receptor, pain signals
Research on	asthma researchers, headache study	Activator (2-1)	headache trigger, headache precipitant
Attribute of clinical study	headache parameter, attack study, headache interview	Inhibitor	adrenoreceptor blockers, influenza prevention
Procedure	tumor marker, genotype diagnosis		
Beginning of activity	headache induction, headache onset	Standard	headache criteria, society standard

Rosario and Hearst relations:

The annotators used 10/38 of the Rosario and Hearst relations to annotate the 101 GO terms (Table 7). The inter-annotator agreement, calculated via Cohen's Kappa, was 0.37 (Table 8). Table 9² shows the confusion matrix for these relations.

Table 8: Overall inter-annotator agreement using Rosario and Hearst relations

TP	FP	FN	P	R	F	κ
71	30	30	0.70	0.70	0.70	0.37

The inter-annotator agreement calculated using the F-measure was 70.29%. The maximum number of terms were annotated as *Activity/Physical Process* followed by *Characteristics* and *Material: Bind*. The annotators observed that there was no good representation for movement terms (for example, *cilium movement*). The inter-annotator agreement is good, given the fact that no training was provided. Again, the disagreements were quite systematic. Of the 30 disagreements, 20 (2/3) were from one cell of the table: Annotator 1 classified 20 compounds as having the *Material: Bind* relation, while Annotator 2 classified the same 20 compounds as having the *Characteristic* relation.

Discussion

The inter-annotator agreement was much higher for the Rosario and Hearst relations than for the Generative Lexicon relations. This is surprising, since the set of Rosario and Hearst relations is much larger than the set of Generative Lexicon relations.

It's premature to say why this is the case, but we can propose some avenues for future investigation: (a) This result might be related to the fact that the Rosario and Hearst relations are domain-specific, while the Generative Lexicon relations are not; (b) this result might be related to the fact that the Generative Lexicon relations are abstract and theoretically motivated, while the Rosario and Hearst relations are concrete and motivated by practical considerations; (c) it might be related to the observation that the annotators only used 10 of the Rosario and Hearst relations implying that the difference in size might not be as big as it seems and the difference in IAA may not be quite as surprising; (d) the difference IAA might go away with actual annotation guidelines and training; and (e) we should also point out that the affordances of the two are different--in particular, the Rosario and Hearst relations might be better for defining information extraction tasks while the Generative Lexicon's relations might be better for supporting inference.

Conclusion

The assumption behind the methodology that was applied here is that inter-judge agreement on annotation task is capable of finding problems in a set of semantic relations. The inter-annotator agreement in the cases of both proposed sets of semantic relations approached that of many completed and published corpus annotation projects, even with very minimal guidelines and no real training. The agreement on this task was fairly high in both cases and disagreements were quite consistent, supporting the basic soundness of the relation-based approach to semantic representation and suggesting that it is not overly subjective. From a methodological perspective, the results suggest that higher levels of agreement and reliability can be reached with some training and refinement of the guidelines. IAA was different between

² Table 9 is on the GitHub Repository:

<https://github.com/KevinBrettonelCohen/SemanticRelationsCompoundNouns.git>

the two sets of relations, suggesting that IAA can differentiate between semantic representations, although a number of possible explanations for those differences should be pursued in future work.

The relatively high IAA suggests that the descriptions and examples of the relations in the Bouillon and Rosario and Hearst papers were easy to follow and that the annotators were able to clearly delineate the relations and the tags in most cases. This is consistent with the claim that they are precise and not overly subjective in their interpretability and applicability. Disagreements between the analysts were quite consistent. This suggests that a higher level of agreement and reliability can be achieved with a little training and refinement of the guidelines. This study should be replicated on a larger scale with proper guidelines and training to achieve a higher level of reliability.

An additional benefit of approaching the evaluation of a set of relations through an annotation task was that we uncovered some shortcomings of the relations. We noted that (a) there is no good representation of movement in the Rosario and Hearst relations, and (b) some of the GO terms were not representable at all with those relations. In the case of the Generative Lexicon relations, we observed frequent confusion between Qualia and Argument (especially *used for* and *aims at*). This suggests that there is a need to clarify the demarcation between the two. A fruitful direction for future work would be to evaluate the nature of any correspondences that might exist between the two sets of relations. The work reported here contributes to the basis for such an effort.

Acknowledgements

We thank NIH grants LM008111 and LM009254 to Lawrence E. Hunter and NSF grant IIS-1207592 to Lawrence E. Hunter and Barbara Grimpe for funding this work.

References

- [1] J. Fan, K. Barker, and B.W. Porter, The knowledge required to interpret noun compounds, in: *IJCAI*, Citeseer, 2003, pp. 1483-1485.
- [2] A. Tribble and S.E. Fahlman, Resolving Noun Compounds with Multi-Use Domain Knowledge, in: *FLAIRS Conference*, 2006, pp. 122-127.
- [3] P. Nakov, On the interpretation of noun compounds: Syntax, semantics, and entailment, *Natural Language Engineering* **19** (2013), 291-330.
- [4] B. Rosario and M. Hearst, Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy, in: *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP-01)*, 2001, pp. 82-90.
- [5] M. Lauer and M. Dras, A probabilistic model of compound nouns, *arXiv preprint cmp-lg/9409003* (1994).
- [6] V. Nastase, P. Nakov, D.O. Seaghdha, and S. Szpakowicz, Semantic relations between nominals, *Synthesis Lectures on Human Language Technologies* **6** (2013), 1-119.
- [7] D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan, and R. Quirk, *Longman grammar of spoken and written English*, MIT Press, 1999.
- [8] N.M. Linh, Noun-noun combinations in technical English, (2010).
- [9] S.D. Joshi and J. Roodbergen, *Patañjali's Vyākaraṇa-Mahābhāṣya Sthānivadhbhāṣika: P. 1.1. 56-1.1. 57*, Bhandarkar Oriental Research Institute, 1990.
- [10] D.O. Séaghdha, Learning compound noun semantics, *University of Cambridge, Cambridge, UK* (2008).
- [11] B. Verhoeven, W. Daelemans, M. Van Zaanen, and G. Van Huyssteen, Automatic Compound Processing: Compound Splitting and Semantic Analysis for Afrikaans and Dutch, *ComAComA 2014* (2014), 20.
- [12] S. Tratz and E. Hovy, A taxonomy, dataset, and classifier for automatic noun compound interpretation, in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2010, pp. 678-687.
- [13] L. Vanderwende, Algorithm for automatic interpretation of noun sequences, in: *Proceedings of the 15th conference on Computational linguistics-Volume 2*, Association for Computational Linguistics, 1994, pp. 782-788.
- [14] C. Dima and E. Hinrichs, Automatic Noun Compound Interpretation using Deep Neural Networks and Word Embeddings, *IWCS 2015* (2015), 173.
- [15] M. Lauer, Conceptual association for compound noun analysis, in: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1994, pp. 337-339.
- [16] A.A. Leenaars and W.D. Balance, A logical empirical approach to the study of suicide notes, *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* **16** (1984), 249.
- [17] P. Bouillon, E. Jezek, C. Melloni, and A. Picton, Annotating qualia relations in Italian and French complex nominals, (2012).
- [18] J. Pustejovsky, The generative lexicon, (1995).
- [19] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, and J.T. Eppig, Gene Ontology: tool for the unification of biology, *Nature genetics* **25** (2000), 25-29.
- [20] G.O. Consortium, Gene ontology consortium: going forward, *Nucleic acids research* **43** (2015), D1049-D1056.
- [21] P.V. Ogren, P.G. Wetzler, and S. Bethard, ClearTK: A UIMA toolkit for statistical natural language processing, *Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP* **32** (2008).
- [22] S. Bethard, P.V. Ogren, and L. Becker, ClearTK 2.0: Design Patterns for Machine Learning in UIMA, in: *LREC*, 2014, pp. 3289-3293.
- [23] J. Cohen, A coefficient of agreement for nominal scales, *Educational and psychological measurement* **20** (1960), 37-46.
- [24] M. Gamer, J. Lemon, I. Fellows, and P. Singh, irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84, Internet resource: <http://CRAN.R-project.org/package=irr>. Verified April 10 (2012), 2013.

Address for correspondence

Kevin Bretonnel Cohen, e-mail id : kevin.cohen@gmail.com

Graph Clustering System for Text-Based Records in a Clinical Pathway

Takanori Yamashita^{a,b}, Naoya Onimura^b, Hidehisa Soejima^c, Naoki Nakashima^a, Sachio Hirokawa^d

^a Medical Information Center, Kyushu University Hospital, Fukuoka, Japan

^b Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

^c Saiseikai Kumamoto Hospital, Kumamoto, Japan

^d Research Institute for Information Technology Kyushu University, Fukuoka, Japan

Abstract

The progressive digitization of medical records has resulted in the accumulation of large amounts of data. Electronic medical data include structured numerical data and unstructured text data. Although text-based medical record processing has been researched, few studies contribute to medical practice. The analysis of unstructured text data can improve medical processes. Hence, this study presents a clustering approach for detecting typical patient's condition from text-based medical record of clinical pathway. In this approach, the sentences in a cluster are merged to generate a "sentence graph" of the cluster after classified feature word by Louvain method. An analysis of real text-based medical records indicates that sentence graphs can represent the medical treatment and patient's condition in a medical process. This method could help the standardization of text-based medical records and the recognition of feature medical processes for improving medical treatment.

Keywords:

Critical Pathways; Data Mining; Mathematical Computing

Introduction

The digitization of medical treatment has resulted in large amounts of medical data to be accumulated. Electronic medical data includes structured numerical data and unstructured text data. Disease diagnosis, medical examination results, and prescriptions are typical examples of structural numerical data that can be used for secondary purposes. In contrast, daily medical records, admission records and discharge summaries are typical examples of unstructured text data. One problem with text data is that there is no unified format for describing the contents. This is, a wide variety of expressions can be used express one idea depending on who writes the contents. However, text-based records include factors that are essential, such as the daily status of patients and medical treatments. Thus, the secondary use of text data is a significant issue.

A clinical pathway in the field of medicine can be thought of as a critical path in engineering. A clinical pathway determines the standard medical procedures for an inpatient with respect to each disease and medical treatment. It clarifies the role of medical staff and provides a schedule. As a result, systematic and high-quality medical treatment can be provided. Furthermore, clinical pathways should improve medical management by providing efficient medical treatment. The Japanese Society for Clinical

Pathway [1] promotes the construction of a standard electronic clinical pathway to standardize medical treatment and improve medical processes. They also encourage the adoption of clinical pathway methods for team medical treatment.

The "all-variance type outcome-oriented clinical pathway" (Figure 1) is a series of medical treatment units that comprise three layers: (a) outcomes, (b) assessments, and (c) tasks. A unit represents a minimal process and is scheduled by combining multiple units. An outcome determines the condition of a patient. Multiple outcomes are the goals of medical staff and a patient for the scheduled duration of hospitalization.

Doctors or nurses in medical practice maintain records of their tasks and assessments of patients' conditions. A variance is recorded in an outcome layer when a patient's condition does not meet the criteria of goal of an assessment layer. When a variance occurs, the staff record the patient's symptoms and future treatment plan as free text in the medical record. This medical record is called a "text-based variance record". Thus, the abnormal condition of the patient and change in the medical intervention plan can be understood within the clinical pathway. The clinical pathway is hence useful for managing medical treatment [2]. In this study, we focus on text-based variance records written in Japanese and apply text mining and visualization for the ultimate goal of treatment standardization.

Several studies on medical record text mining have been conducted. They are mainly used for classification and performance evaluation. However, only few studies have been conducted wherein specific sentences and words are used to extracted patients' conditions [3-5]. An analysis of unstructured free text could potentially extract a large amount of information regarding patient health status and medical histories [6;7]. Although feature words extracted using machine learning and feature selection can be evaluated, a patient's condition is not extracted from the text-based record [8;9]. Applying text mining to Japanese is difficult because Japanese is not easy to judge the boundaries of words.

The present study proposes a visualization system for text-based Japanese records of clinical pathway so that an overall picture of the variances can be understood. We developed a system that classifies the sentences in the variances and generates a "sentence graph" or template that describes each cluster. We confirmed that the feature words in a sentence graph correspond to the tasks and assessments in the pathway data. This visualization method can be used as a guide for the standardization of variance description.

Related Work

In order to extract information from a large set of documents, search engines have become indispensable. In some search task, the user is satisfied by only one item of information. However, the situation is different when search is performed to gain an overall understanding of a topic such as research trends and or a reputation. For such task, there are two approaches: document summary and key sentence extraction. For the scientific articles, reputation information and FAQ documents, it is known that utilizing characteristic phrases and words in a field improves extraction performance [10]. Alternatively, a method to extract phrases has also been proposed [11]. Important word and sentence extraction is a major theme of text mining. Radev et al. [12] applied sentence clustering using feature words and extracted the centroids of each cluster as the important sentences. Mihalcea and Tarau [13] proposed the TextRank method, which considers the network of keywords and uses a HITS-like algorithm [14] to evaluate keyword importance. Pantel and Pennacchiotti [15] considered the problem from a different viewpoint and proposed the Espresso-type bootstrapping method to extract information from documents.

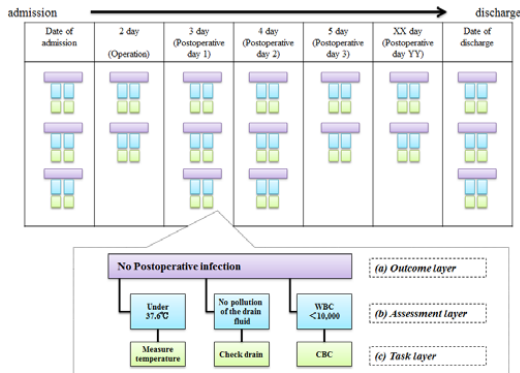


Figure 1 – All-variance type outcome-oriented clinical pathway (three layer)

These methods were developed independently. However, they can be considered within the same framework of word and document co-occurrence graphs analysis [16;17]. Both approaches extract the important keywords and sentences as tightly connected sub-graphs of a bipartite graph that comprises sentences and words as their nodes and the edges between them.

The text-based variance records analyzed in this study are very short sentences of 18 characters on average. Simple key sentence extraction is not suitable for these sentences. We need something similar to a template comprising similar multiple sentences. Hence, we first composed a bipartite graph of sentence and word nodes. We then partitioned the graph into a set of dense sub-graphs and obtained a set of clusters. Finally, we constructed the “sentence graph” by merging the words using the Louvain method [19;20], an improved version of the Newman method.

Data and Methods

Text-based variance records

In this study, we analyzed text-based variance records of inpatients from the clinical pathways of Kumamoto-Saiseikai Hospital from April 2014 to April 2016 (Figure 2).

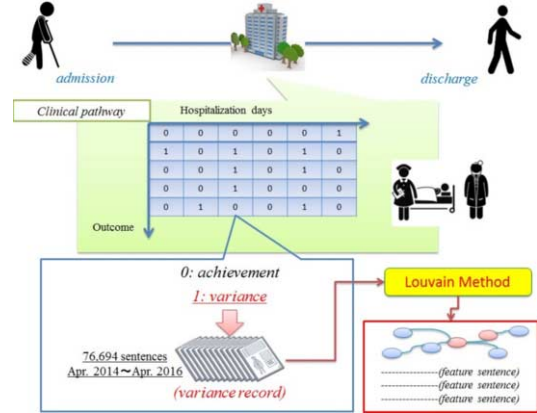


Figure 2 – Visualization concept

We collected 163,004 sentences from text-based variance records written in Japanese. We then applied morphological analysis to the sentences and obtained 174,145 words. In this study, we focused on 756 words that represent more than 100 of the sentences to obtain the common templates. In addition, we considered the 98,004 sentences that contain two or more of these words.

Clustering of sentences

First, we applied the Louvain method [19;20] to cluster the sentences. For each cluster, we then constructed a graph whose nodes are either sentences or words and whose edges are pairs comprising a sentence and a word. An edge of the graph is a pair (s_i, w_j) of sentence s_i and word w_j , where w_j occurs in s_i . Further we partitioned the graph using the Louvain method to obtain 532 dense sub-graphs: C_1, C_2, \dots, C_{532} . Table 1 shows the top 20 clusters according to the number of sentences. For words of C_{18} and C_{19} , although it is a different word in Japanese, it is a word of the same meaning in English (Table 1).

Merging sentences

For the common template of each cluster, we constructed the sentence graph $G(C_k)$ of cluster $C_k = \{s_1, \dots, s_m\}$ using the following procedure. First, we transformed a sentence s_i in cluster C_k into a sequence $seq(s_i) = (w_1, w_2, \dots, w_n)$ of words, where $s_i = w_1 \dots w_n$. Next, we created a line graph $G(s_i)$ with an edge in the form of (w_j, w_{j+1}) for $j=1, \dots, n-1$ and obtained line graphs $G(s_1), \dots, G(s_m)$. We merged the nodes with the word of their labels and merged edges that have the same start and end nodes. Finally, we discarded the edges with a frequency of less than 20% of the most frequent edges and selected a node associated with the top edge to obtain $G(C_k)$.

Results and Discussion

Sentence graphs

The sentence graphs were generated by the above method represents typical medical treatment described in text-based. Figure 3 shows the sentence graphs of clusters C_1 , C_3 , C_7 , C_{17} , and C_{20} . Each line in a graph represents a sentence in the cluster. The word of red colored node belong to the cluster. The numerical value in the node represents the number of words, and the numeral on the edge represents the number of combinations between the nodes. For example, we can read the path in the sentence graph of cluster C_3 as “bradycardia / symptom / no / because / observe the progress / to (particle) / suru (verb)”. Note that the words “bradycardia”, “symptom” and “observe the progress” are the feature words of cluster C_3 in Table 1. In C_7 , we can read as “amezinium / effect / poor or effective / observe the progress”. Note that the words “amezinium”, “effect” and “poor” are the feature words of cluster C_7 in Table 1.

The system displays the sentence graphs as well as the frequent sentences of each cluster. The following sentences are typical frequent sentences extracted by the system:

- “To observe (careful, confirmation) circulation dynamics, swallowing disorder, dysarthria, paralysis, respiratory condition”
- “Observe the progress because there are no bradycardias symptom”
- “Urine volume normal, observe the progress of the condition”
- “Observe the progress of the condition with amezinium effect”
- “Observe the progress of the condition after transfusion”
- “Observe the progress of the condition because diastolic blood pressure is not high”
- “Defecation yesterday; observe the progress of the condition”
- “Observe the progress of the condition after administering analgesic (suppository)”
- “Follow on fever type”

The visualization of “sentence graph” helps to interpret the typical sentences from a cluster. Because these sentences are described when variances occur, they have the same pattern “observe the progress of the condition”. In particular, many words were represented with many sentences, and it was not the same content classification in C_1 . It is necessary to narrow the target.

Characteristic task and assessment words

An interesting point about these sentence graphs is that most of the feature words that appear in the graphs are either some tasks that doctors or nurses should do later or that have been done previously, or some assessment that they must observe. Thus, the sentence graphs capture the lower clinical pathway layers of task and assessment.

Furthermore, we found the following two kinds of words and phrases in the sentence graphs after or before the term “observe progress”:

1. “Administering analgesic (suppository)”, “killing pain”, “transfusion” and “follow on fever type”.
2. “There are no bradycardias symptom”, “defecation”, “urine volume normal”, “amezinium effect” and “blood pressure id not high”.

The phrases in the first list describe the tasks of medical treatment. The phrases in the second list are the assessments or observations of medical treatment. Thus, the extracted keywords and phrases capture the feature words of the task and assessment layers in the clinical pathway.

Table 1 – Top 20 clusters in term of the number of sentences *(Japanese word)

Cluster	Sentence count	Word count	Representative feature word
C1	15,680	67	circulation dynamics, swallowing disorder, dysarthria, paralysis, respiratory condition, drain, careful, confirmation, observation...
C2	6,104	26	suppository, analgesics, killing pain, gastric tube, ...
C3	4,714	13	bradycardia, symptom, observe the progress, vital, ...
C4	3,123	27	systolic blood pressure, dialysis, ultrafiltration, ...
C5	3,047	15	dietary intake, urine volume, ...
C6	2,634	5	internal, analgesics, ...
C7	1,468	5	effect, poor, amezinium, ...
C8	1,305	2	to do (<i>si, you</i>)
C9	1,250	4	analgesics, add, unnecessary
C10	1,221	5	blood pressure, low
C11	1,175	6	primary doctor, report, accept instruction, ...
C12	1,061	12	pain, endure, strong, nursing call, ...
C13	1,035	2	follow, fever type
C14	918	3	main, administering
C15	874	4	be considered, ...
C16	776	6	push, pump, ...
C17	746	3	high blood pressure, diastolic
C18	739	2	observe the progress (<i>keika miru</i>)
C19	638	2	observe the progress (<i>yousu miru</i>)
C20	604	2	yesterday, defecation

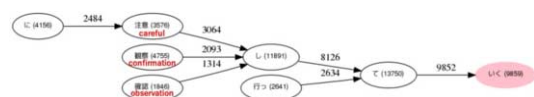
Conclusion and Future Work

This study reports a visualization system for text-based variance records that describe a patient’s condition using free- style text. We applied the Louvain method to cluster the variance records and words and constructed sentence graphs of each cluster of text-based variance records. The sentence graphs display typical task and assessment patterns.

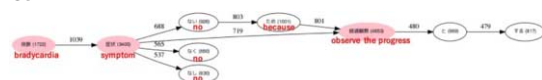
Previous studies on medical documents have focused mainly on the words in the documents [8-9]. However, it is difficult to interpret a patient’s condition from documents using words alone. This study demonstrated a new approach that uses sentences and sentence graphs. The text-based variance records contain valuable information regarding the patient because they are written by medical staff. One of the challenges to making good use of these records is that they have no structure. We found that the sentence graphs and feature words in the graph correspond to the task and

assessment layer of a three-layer clinical pathway. However, it is necessary to confirm the validity because this system expresses only by the word frequency. Ultimately, the sentence graphs and their templates can be adapted into templates that can be filled in by medical staff. The input time and diversity of text expression will be reduced, which will also improve mechanical analysis.

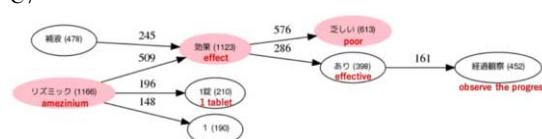
C1



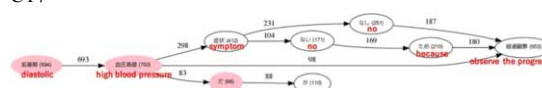
C3



C7



C17



C20

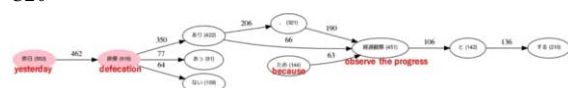


Figure 3 – Sentence graph pattern by using clustering

We plan to evaluate the performance of the proposed method. This analytical method will then be extended to other text-based medical records or focus on clinical cases in further research.

Acknowledgements

This research is partially supported by JSPS KAKENHI Grant Number 15H02778 and 16mk0101064h0001 from Japan Agency for Medical Research and development, AMED.

References

- [1] Japanese Society for Clinical Pathway, <http://www.jscp.gr.jp>
- [2] N. Nakashima, H. Okada, A. and Gouchi, Hospital Information System and Clinical Pathways, *Japan journal of medical informatics* **27** (2007), 21-28. (in Japanese)
- [3] S.V. Pakhomov, P.L. Hanson, S.S. Bjornsen, S.A. Smith, Automatic classification of foot examination findings using clinical notes and machine learning, *J Am Med Inform Assoc* **15** (2007), 198-202.
- [4] D. Mowery, J. Wiebe, S. Visweswaran, H. Harkema, W.W. Chapman, Building an automated SOAP classifier for emergency department reports, *J Biomed Inform* **45** (2012), 71-81.
- [5] H. Kubo, T. Yamashita, Y. Wakata, N. Nakashima and S. Hirokawa, Effect of Synonym on Prediction Performance for Postoperative Hospital Stay by Text Mining, *Proc. AROB 2016*, (2015), 409-413.
- [6] Y. Wang, Z. Yua, Y. Jiangb, Y. Liuc, L. Chena, and Y. Liua, A framework and its empirical study of automatic diagnosis of traditional chinese medicine utilizing raw free-text clinical records, *J Biomed Inform* **45** (2012), 210-223.
- [7] A. Wright, A. B. McCoy, S. Henkin, A. Kale, and D. F. Sittig, Use of a support vector machine for categorizing free-text notes:

Assessment of accuracy across two institutions, *J Amer Med Inf Assoc* **20** (2013), 887-890.

- [8] C. Soguero-Ruiz, K. Hindberg, J.L. Rojo-Alvarez, S.O. Skrovseth, F. Godtliebsen, K. Mortensen, A. Revhaug, R.O. Lindsetmo, K.M. Augestad and R. Jenssen, Support Vector Feature Selection for Early Detection of Anastomosis Leakage From Bag-of-Words in Electronic Health Records, *IEEE J Biomed Health Inform* **20** (2016), 1404-1415.
- [9] T. Yamashita, Y. Wakata, S. Hamai, Y. Nakashima, Y. Iwamoto, B. Flanagan, N. Nakashima and S. Hirokawa, Presumption Model for Postoperative Hospital Days from Operation Records, *International Journal of Computer & Information Science* **16** (2015), 50-59.
- [10] R. Higashinaka and H. Isozaki, Automatically Acquiring Causal Expression Patterns from Relation-annotated Corpora to Improve Question Answering for why Questions, *ACM Trans. on Asian Language Information Processing, Special Issue on NTCIR-6 7* (2008), 29.
- [11] M. Krapivin, A. Autayeu, M. Marchese, E. Blanzieri, and N. Segata, Keyphrases extraction from scientific documents: improving machine learning approaches with natural language processing, in: *International Conference on Asian Digital Libraries*, Springer, 2010, pp. 102-111.
- [12] D.R. Radev, H. Jing, M. Stys and D. Tam, Centroid-based summarization of multiple documents, *Information Processing and Management* **40** (2004), 919-938.
- [13] R. Mihalcea and P. Tarau, TextRank: Bringing Order into Texts, *In Proceedings of EMNLP 2004*, (2004), 404-411.
- [14] G. Palshikar, Keyword extraction from a single document using centrality measures, *Pattern recognition and machine intelligence* (2007), 503-510.
- [15] P. Pantel and M. Pennacchiotti, Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations, *Proc. 44th ACL*, (2006), 113-120.
- [16] M. Komachi, T. Kudo, M. Shimbo and Y. Matsumoto, Graph-based Analysis of Semantic Drift in Espresso-like Bootstrapping Algorithms, *Proc. EMNLP 2008*, (2008), 1011-1020.
- [17] D. Radev and R. Mihalcea, Network and Natural Language Processing, *AI Magazine*, (2008), 16-28.
- [18] M.E.J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* **69** (2004).
- [19] V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, Fast unfolding of communities in large networks, *Journal of statistical mechanics: theory and experiment* **10** (2008), P10008.
- [20] Louvain method for community detection in large graphs, <https://sourceforge.net/projects/louvain/>

Address for correspondence

Takanori YAMASHITA, t-yama@info.med.kyushu-u.ac.jp

Research on Ratio of Dosage of Drugs in Traditional Chinese Prescriptions by Data Mining

Yu Xing-wen, Gong Qing-yue, Hu Kong-fa, Mao Wen-jing, Zhang Wei-ming

College of Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

Abstract

Maximizing the effectiveness of prescriptions and minimizing adverse effects of drugs is a key component of the health care of patients. In the practice of traditional Chinese medicine (TCM), it is important to provide clinicians a reference for dosing of prescribed drugs. The traditional Cheng-Church biclustering algorithm (CC) is optimized and the data of TCM prescription dose is analyzed by using the optimization algorithm. Based on an analysis of 212 prescriptions related to TCM treatment of kidney diseases, the study generated 87 prescription dose quantum matrices and each sub-matrix represents the referential value of the doses of drugs in different recipes. The optimized CC algorithm can effectively eliminate the interference of zero in the original dose matrix of TCM prescriptions and avoid zero appearing in output sub-matrix. This results in the ability to effectively analyze the reference value of drugs in different prescriptions related to kidney diseases, so as to provide valuable reference for clinicians to use drugs rationally.

Keywords:

Medicine, Chinese Traditional; Prescriptions; Data Mining

Introduction

Biclustering algorithms is a field of research that is being developed and its algorithm is widely used in gene expression data analysis [1-3], literature research hotspot analysis [4-6] and biological data analysis [7-8]. Using double clustering algorithms can obtain doses in different traditional Chinese medicine TCM prescriptions, elaborate scientifically dose-effect relationship, give a reasonable choice of TCM dose, and lay the foundation for the safe and effective medication [9].

Current biclustering algorithms can be divided into traditional clustering, iterated greedy search biclustering, exhaustive enumeration biclustering and mathematical modeling biclustering [10]. Among them, traditional clustering, such as coupled two-way clustering (CTWC) based on hierarchical clustering [11], cannot avoid the influence of global clustering nor can-generate favorable local submatrices, despite of their simple realization. This paper adopts the improved CC algorithm based on the iterated greedy search for the prescription dose analysis. The computing speed of the algorithm is faster than the exhaustive enumeration biclustering [12], and can effectively avoid global disturbance to obtain valid biclustering submatrices.

Methods

CC biclustering algorithm

Being one of the earliest biclustering algorithms, Cheng-Church biclustering algorithm [13] was first proposed by Cheng and Church in 2001. Sticking to the principle of iterated greedy search, this algorithm can generate a submatrix meeting the regulated threshold value in each search.

Mathematical description of CC algorithm

CC biclustering algorithm introduces the concept of mean squared residue to describe degree of internal similarity or consistency of submatrices. $H(I, J)$ refers to a mean squared residue of the submatrix whose row number is $|I|$ and column number is $|J|$.

$$H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ij} - a_{i.} - a_{.j} + a_{..})^2 = \frac{1}{|I||J|} \sum_{i \in I, j \in J} RS_{ij}^2$$

$$a_{i.} = \frac{1}{|J|} \sum_{j \in J} a_{ij}$$

$$a_{.j} = \frac{1}{|I|} \sum_{i \in I} a_{ij}$$

$$a_{..} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} a_{ij}$$

$$RS_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..}$$

In a submatrix, RS_{ij} is a residue; $a_{i.}$ is the mean of Row i , $a_{.j}$ is the mean of Column j ; $a_{..}$ is the mean of the submatrix. Where, the larger the value of $H(I, J)$ is, the smaller the degree of similarity within the submatrix is. The smaller the value of $H(I, J)$ is, the more consistent the internal value of the submatrix is. Set ε to be maximal threshold value of the mean squared residue. All matrices satisfying the condition of $H(I, J) \leq \varepsilon$ is adopted as the target output submatrices of CC biclustering algorithm.

To find a submatrix satisfying the condition of $H(I, J) \leq \varepsilon$ in one search. First, the target matrix should undergo row and column deletion. When the mean squared residue of certain row (column) of the submatrix is larger than ε , then:

$$R = \{i \in I; \frac{1}{|J|} \sum_{j \in J} RS_{ij}^2 > H(I, J)\}$$

Then, the row (column) is deleted to efficiently reduce the value of its mean squared residue. During the whole column and row deletion process, every iteration deletes the maximum column or row with the maximal mean squared residue until the mean squared residue of the submatrix is smaller than the threshold value, ε . Then, the qualified matrix can be preliminarily obtained:

Since the submatrix obtained by deletion of columns and rows is not the maximal submatrix, it is necessary to increase rows and columns of the submatrix:

If the average of mean squared residue-of the column (row) within the submatrix is smaller than the threshold value, ϵ , then:

$$R=\{i \notin I; \frac{1}{|J|} \sum_{j \in J} RS_{ij}(i, j)^2 \leq H(I, J)\}$$

Then, substitute the column (row) into the submatrix. The mean squared residue of the newly-generated submatrix is smaller than or equal to ϵ , so qualified columns and rows are added to generate the maximal submatrix satisfying the output condition. After the qualified submatrix is output, the random number substitution is conducted before the next round of searching. To be specific, a random number matrix whose size is similar to the submatrix is used to substitute the submatrix to destroy the original consistency of the matrix and to obtain different qualified submatrices in every search.

Applications of CC biclustering algorithm to prescription dose analysis

The research made the dose of the prescription dose into the dose of the original matrix , according to the horizontal axis name and the vertical axis square name. Each element in the matrix corresponds to the dose of a drug in a prescription , as shown in Table 1.

Table 2 is the use of traditional CC algorithm for traditional Chinese medicine prescription data on the double-cluster analysis to output some sub-matrix. The analysis set the random number range out of the prescription dose range (0-500g), excluding the effect of the random number on the resulting sub-matrices. In the analysis results, the sub-matrix contains more 0 values, and the number of iterations increases with the algorithm. The proportion of 0 values gradually increased, seriously affecting the generation of effective results. The reason for the above results is mainly due to the unique nature of TCM prescription dose analysis and the limitations of ordinary CC algorithm, and the specific existent problems are the following two aspects:

Table 1 – Prescription dose original matrix

Prescription	Ginseng				
	Angelica	g	Fluorite	Authentic colla	corii asini
Radix Morindae Officinalis	0	22.5	0		
Dragon Tiger Panacea	15	0	0		0
Ten Supplements Pill	60	60	0		0
Jiaotai Pill	0	0	0		0
Dendrobium Powder	0	0	0		0
Guoxia Pill	0	0	0		0

Prescription	Fangfeng root		Chuanxiong		Angelica
	leaf	Cnidium	Digupi		
Keel pill	1.5	0	0	0	0
Eyesight improving water pill	0	0	0	0	0
31 Shenqi Pill	0	0	0	0	0
Kanlijiji Pill	0	0	0	0	0
Suppressing yang wine	0.9	0	0	0	0.9

Prescription	Chinese Photinia leaf		Chinese yam		Ceylon
	Chrysanthemum		Huaisheng Rehmannia		
Luminous pill	0	0	0	0	0
Tianxiong pill	0	0	0	0	0
Rehmannia pill	0	0	0	0	0

Guben-jianyang-Pill	0	0	0	0	0
Lysol soup	0	0	0	0	0

First, since the number of drugs in every prescription is smaller than the total number of drugs in all prescriptions, 0 appears in many places of the matrix. The submatrix composed of 0 is exactly the output submatrix of CC biclustering algorithm, so the output result using the traditional CC biclustering algorithm is mostly a submatrix composed of 0. However, 0 does not make any sense for the TCM dose analysis. Worse still, it might influence validity of the target submatrix.

Second, in each search, CC biclustering algorithm should conduct the random value substitution. In the traditional CC biclustering analysis, though these random numbers might impose certain inference on the submatrix output, the results obtained are still reliable. However, in the prescription dose matrix, the original matrix has too many 0s. The interference of the random number might seriously influence the output of valid submatrices. Worse still, the random number submatrix might be output, thus seriously influencing reliability of results.

To sum up, the key to optimizing biclustering algorithms is to exclude the interference of 0 on the prescription dose matrix and to seek a method more effective than the random number substitution method.

Optimization of CC algorithm

Improvement of CC in this paper refers to the improvement strategy of the POBA [14]. The punishment strategy is introduced to eliminate 0 from the matrix during the calculation process.

First, a punishment matrix, W, with a size similar to that of the prescription dose matrix is built. Let the initial value of the matrix element be 0. If the original prescription dose matrix element is 0, let the element value at the same position of the punishment matrix be 1.

Then, the calculation process of the mean squared residue by CC biclustering algorithm is optimized:

$$H(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot})^2 + (W_{ij} * \theta)$$

The above equation adds the “punishment formula” to the original calculation equation of the mean squared residue. W_{ij} stands for the element of the punishment matrix in Row i and Column j; θ controls the degree of punishment. In practical applications, 0 makes no sense to TCM prescription dose analysis. On the contrary, it might influence the output of the target submatrix. Therefore, in order to guarantee full cleaning of 0 in the output submatrix, let θ be:

$$\theta = |I| * |J| * \epsilon$$

Then, there will contain no 0 in the output submatrix. In the following part, it is necessary to improve the column and row deletion and adding process of the submatrix so as to guarantee smooth generation of the target submatrix:

$$R=\{i \in I; \frac{1}{|J|} \sum_{j \in J} RS_{ij}(i, j)^2 + (W_{ij} * \theta) > H(I, J)\}$$

The above equation can delete columns or rows containing more 0s in advance during the column and row deletion process.

$$R=\{i \notin I; \frac{1}{|J|} \sum_{j \in J} RS_{ij}(i, j)^2 + (W_{ij} * \theta) \leq H(I, J)\}$$

The above equation can guarantee adding of columns or rows not containing 0s during the column and row adding process.

In practical applications, let θ be:

$$\theta = |I| * |J| * \varepsilon$$

Finally, after the end of every search, let the element value of the punishment matrix whose position is the same to the output submatrix be 1. Replacing the random number substitution method with this method can guarantee full cleaning of 0s in the output submatrix and satisfy demands of outputting valid target submatrices.

Results

An original matrix was built for relevant TCM prescriptions to treat kidney diseases. The improved CC algorithm is used for biclustering analysis. The prescription dose submatrix is output to obtain referential value of different drugs in different prescriptions.

Building of the prescription dose original matrix

All prescriptions are from TCM Prescription Dictionary[6] with “stranguria”, “impotence”, “turbid semen”, “kidney essence deficiency” and “retention of urine” as search terms. In total, 212 TCM prescriptions to treat kidney diseases are screened out.

The prescription data are further screened. With the prescription name as the row and the drug name as the column, the prescription dose original matrix is built. This is shown in Table 1 above.

Generation of the prescription dose submatrix

Using the improved CC clustering algorithm and setting the iterations to be 120 and ε be 10, the prescription dose original matrix outputs 87 prescription dose submatrices in total. Every submatrix stands for the mean dose of different drugs in different prescriptions. This is shown in Table 2 below:

Table 2 – Prescription dose submatrix

Prescription name	Schisandra chinensis	Cistanche	Fructus lyci
Yeguang Pill	15	15	23
Shihu Mingmu Pill	12.5	12.5	10
Gorgon Euryale Pill	30	30	30

Prescription name	White poria	Astragalus	Radix sileris
Longgu Pill	2.1	2.1	1.5
Rabbit Liver Pill	23	30	23

Prescription name	Shaved cinnamon bark	Notopterygium incisum
Radix Achyranthis Bidentatae Pill	22.5	30
Tianxiong Powder	30	22.5

Discussion

Take one last submatrix as an example. Radix Achyranthis Bidentatae Pill and Tianxiong Powder contain shaved cinnamon bark and notopterygium incisum. In the former prescription, the dose of shaved cinnamon bark and the notopterygium incisum is 22.5g and 30g, respectively. The ratio between the two is 3:4. This prescription highlights the function of notopterygium incisum to relieve pains and treat rheumatism. In the latter prescription, the dose of shaved cinnamon bark and notopterygium incisum is 30g and 22.5g. The ratio between the two is 4:3. This prescription highlights the function of shaved cinnamon bark to treat wind-pathogenic kidney diseases. Thus, different doses of different drugs in a

prescription follow certain rule. The flexibility of TCM prescriptions lies in its dose. Different doses focus on different treatment functions. Even if the same drugs are included in different prescriptions, different doses can result in different functions. In the clinical medication, the doctor should be based on the actual situation of each patient, according to the results of the analysis, and flexibly choose a different dose ratio to achieve different therapeutic purposes.

The iterated greedy search has its inherent defect, which is prone to get the local optimal solution. Even the optimized CC biclustering algorithm in this paper cannot get rid of the limitation. It is impossible to output high-quality submatrix in every iteration. Thus, it is necessary to increase iterations and further filter the final output submatrix using programming techniques.

However, the optimized CC algorithm can effectively eliminate the interference of 0s in the TCM dose original matrix, thus avoiding outputting meaningless submatrices containing 0s. This can greatly promote applicability of biclustering algorithm to TCM dose analysis. Meanwhile, without using the random number substitution method, the algorithm proposed in this paper can avoid negative influence of the random number on the output submatrices obtained through iterations.

Conclusion

The prescription dose ratio is an important part of the TCM prescription compatibility. A drug can have different doses in different prescriptions, thus playing a different role. Drug dose forms the soul of the prescription compatibility. Moderateness of drug doses can directly influence functions and clinical effects of prescriptions, and even life safety of patients. ZHANG Jiebin (1563~1640), a doctor of the Ming Dynasty, said that, “To create a disease, drug dose should not be too little nor too much. If being too little, the drug dose cannot help treat a disease; if being too much, the drug dose might cause other injuries and diseases.” Therefore, doctors should attach great importance to the influence of drug dose on clinical prescriptions under the prerequisite of sticking to the principle of “Jun-Chen-Zuo-Shi”. In this study, the improved biclustering algorithm was used to analyze the dose reference values of different drugs associated with nephropathy in different prescriptions. It can help clinicians optimize prescription, maximize the efficacy of prescriptions, and minimize the adverse effects of drugs on patients. To furtherly guide the physician to use drugs rationally based on the analysis results, combine with all aspects of environmental factors and the expected treatment results, flexibly choose a different dose ratio, and achieve optimal treatment.

Acknowledgements

This study was supported by The National Natural Science Foundation (Project number: 81674099).

References

- [1] Williams, A. and S. Halappanavar, *Application of biclustering of gene expression data and gene set enrichment analysis methods to identify potentially disease causing nanomaterials* [J]. Beilstein J Nanotechnol, 2015. 6: 2438-2448.
- [2] Gonzalez-Calabozo, J.M., F.J. Valverde-Albacete and C. Pelaez-Moreno, *Interactive knowledge discovery and data mining on genomic expression data with numeric formal concept analysis*. BMC Bioinformatics [J], 2016. 17(1): 374.
- [3] Fu R.Y., Huang J., Hu B.Q., Pang C.Y., *Application of Alzheimer's Gene Expression Data to Compare Two Kinds of Hierarchical Clustering Methods* [J]. Journal of Sichuan Normal University (Natural Science Edition), 2015, (06): 925-929.

- [4] Li, F., et al., *Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis*. *J Med Internet Res* [J], 2015. 17(3): e81.
- [5] Jiang J.Z., *Literature Research Hotspot analysis on Chinese medicine treatment of cerebral infarction* [J]. Science and Technology Information Development and Economy, 2015, (17): 140-142 + 160.
- [6] Gong X.C., Zhao Y.G., An X.Y., *Study on the frontier of breast cancer related enzymes based on double clustering method* [J]. Chinese Journal of Medical Library and Information Science, 2016, (02): 69-74.
- [7] Backman, T.W., D.S. Evans and T. Girke, *Large-scale bioactivity analysis of the small-molecule assayed proteome* [J]. PLoS One, 2017. 12(2): e0171413.
- [8] Zhao Y.N., Wei Z., Wu X.L., Guo Y.Q., Sun C., Zhong S.N., Zhao C.Y., *Double Clustering Analysis Of Drug Resistance Of Gram-Negative Bacteria* [J]. Chinese Journal Of Microecology, 2015, (05): 531-533
- [9] Wang R. X. *Prescription Dose Pattern Research Based On Bicluster Algorithm* [J]. Liaoning Journal of Traditional Chinese Medicine, 2016, 01: 8-9.
- [10] Zhang M. & Ge W. H. *The Research And Advances On Biclustering* [J]. Microcomputer & Its Applications, 2012, 04: 4-6+10.
- [11] Getz G, Levine E, Domany E. *Coupled Two-Way Clustering Analysis Of Gene Microarray Data* [C]. In Proceedings Of The Natural Academy Of Sciences Usa, 2000: 12079-12084.
- [12] Tanay A, Sharan R, Et Al. *Revealing Modularity And Organization In The Yeast Molecular Network By Integrated Analysis Of Highly Heterogeneous Genomewide Data* [C]. Proc Natl Acad Sci U S A., 2004: 2981-6.
- [13] Cheng Y, Church Gm. *Biclustering of Expression Data* [C]. Proc. Eighth Int'l Conf. Intelligent Systems for Molecular Biology (Ismb'00), 2000: 93-103.
- [14] Zhou Ch. *Research And Applications Of Bicluster Algorithm Based On High-Dimensional Data* [D]. Nanjing University of Science and Technology, 2009.
- [15] Peng H. R. *Tcm Prescription Dictionary* [M]. People's Medical Publishing House Co., Ltd., 1996.

Addresses for correspondence

Gong Qing-yue (corresponding author): qygong@126.com

Yu Xing-wen: 437472167@qq.com

Retrofitting Concept Vector Representations of Medical Concepts to Improve Estimates of Semantic Similarity and Relatedness

Zhiguo Yu^a, Byron C. Wallace^b, Todd Johnson^a, Trevor Cohen^a

^a The University of Texas School of Biomedical Informatics at Houston, Houston, Texas, USA

^b College of Computer and Information Science, Northeastern University, Boston, Massachusetts, USA

Abstract

Estimation of semantic similarity and relatedness between biomedical concepts has utility for many informatics applications. Automated methods fall into two categories: methods based on distributional statistics drawn from text corpora, and methods using the structure of existing knowledge resources. Methods in the former category disregard taxonomic structure, while those in the latter fail to consider semantically relevant empirical information. In this paper, we present a method that retrofits distributional context vector representations of biomedical concepts using structural information from the UMLS Metathesaurus, such that the similarity between vector representations of linked concepts is augmented. We evaluated it on the UMNSRS benchmark. Our results demonstrate that retrofitting of concept vector representations leads to better correlation with human raters for both similarity and relatedness, surpassing the best results reported to date. They also demonstrate a clear improvement in performance on this reference standard for retrofitted vector representations, as compared to those without retrofitting.

Keywords: Semantics; Natural Language Processing; Unified Medical Language System

Introduction

Incorporation of semantically related terms and concepts can improve the retrieval [1; 2] and clustering [3] of biomedical documents; enhance literature-based discovery [4; 5]; and support the development of biomedical terminologies and ontologies [6]. However, automated estimation of the semantic relatedness between medical terms in a manner consistent with human judgment remains a challenge in the biomedical domain. Many existing semantic relatedness measures leverage the structure of an ontology or taxonomy (e.g. WordNet, the Unified Medical Language System (UMLS), or the Medical Subject Headings (MeSH)) to calculate, for example, the shortest path between concept nodes [7-9]. Alternatively, vector representations derived from distributional statistics drawn from a corpus of text can be used to calculate the relatedness between concepts [7; 10]. Other corpus-based methods use information content (IC) to estimate the semantic relatedness between two concepts, from the probability of these concepts co-occurring [9; 11; 12]. This raises the question of whether knowledge- or corpus-based metrics are most consistent with human judgment.

In 2012, Garla and Brant [13] evaluated a wide range of lexical semantic measures, including both knowledge-based approaches leveraging the structure of an ontology or taxonomy [7; 14; 15] and distributional (corpus-based)

approaches relying on co-occurrence statistics to estimate relatedness between concepts [16; 17]. This systematic investigation used several publicly available benchmarks. The most comprehensive of these is the University of Minnesota Semantic Relatedness Standards (UMNSRS), which contains the largest number and diversity of medical term pairs of any reference standard to date [18]. Medical terms in the set have been mapped to Concept Unique Identifiers (CUIs) in the UMLS, and term pairs have been annotated by human raters for similarity (e.g. *Lipitor* and *Zocor* are similar) and relatedness (e.g. *Diabetes* and *Insulin* are related). The best Spearman rank correlation for relatedness and similarity on this benchmark reported in [13] are 0.39 and 0.46 respectively.

Neural network based models that are trained to predict neighboring terms to observed terms, such as the architectures implemented by the *word2vec* package [19], have gained popularity as a way to obtain distributional vector representations of terms. Vectors induced in this way have been shown to effectively capture analogical relationships between words [20], and under optimized hyperparameter settings these models have been shown to achieve better correlation with human judgment than prior distributional models such as Pointwise Mutual Information (PMI) and Latent Semantic Analysis (LSA) on some word similarity and analogy reference datasets [21; 22]. However, embedding models are trained on terms, not concepts. In 2014 De Vine [23] and his colleagues demonstrated that word embedding models trained on sequences of UMLS concepts (rather than sequences of terms) outperformed established corpus-based approaches such as Random Indexing [24] and LSA [25].

In 2014 Sajadi et al. reported that a graph-based approach (HITS-sim) leveraging Wikipedia as a network outperformed *word2vec* trained on the OHSUMED corpus for the UMNSRS benchmark, with Spearman rank correlations of 0.51 and 0.58 for semantic relatedness and similarity respectively [26]. Most recently, Pakhomov et al. [27] performed an evaluation of *word2vec* trained on text corpora in different domains - Clinical Notes, PubMed Central (PMC), and Wikipedia - and achieved higher correlations of 0.58 and 0.62 for semantic relatedness and similarity respectively, which are the best results reported to date on the UMNSRS benchmark.

However, while vector representations produced by neural word embedding models are semantically informative, they disregard the potentially valuable information contained in semantic lexicons such as WordNet, FrameNet, and the Paraphrase Database. In 2015, Faruqui et al. developed a 'retrofitting' method that addresses this limitation by incorporating information from such semantic lexicons into word vector representations, such that semantically linked words will have similar vector representations [28]. In our previous work, we have tested this approach as a way to

improve measures of semantic relatedness between MeSH terms using information from the MeSH taxonomic structure [29]. While retrofitting word vectors resulted in higher correlation with physician judgments, the reference set utilized was the MiniMayoSRS benchmark [7], which is a relatively small dataset (29 medical concept pairs). Furthermore, we did not apply neural word embeddings, which have been shown to outperform prior distributional models on this task.

In this paper, we extend our previous ‘retrofitting’ work in the following ways: (1) We use one of *word2vec*’s models to construct vector representations; (2) For construction of vector representations of UMLS concepts, we follow the approach described in [23] and train our model on sequences of UMLS medical concepts extracted from all of MEDLINE’s titles and abstracts; (3) We evaluate our approach with a more extensive reference standard, the UMNSRS benchmark. Our results show that our method achieves higher correlation with human ratings for relatedness and similarity than the best results reported so far on UMNSRS benchmark [27].

Methods

Reference Standard

We used the University of Minnesota Semantic Relatedness Standard (UMNSRS) as our evaluation data [18]. This dataset consists of over 550 pairs of medical terms. Each term has been mapped to a CUI in the UMLS. Each pair of terms was assessed by 4 medical residents and scored with respect to the degree to which the terms were similar or related to each other, using a continuous scale. There are two subsets in UMNSRS - UMNSRS-Similarity and UMNSRS-Relatedness. UMNSRS-Similarity contains 566 pairs of terms rated by 4 medical residents. UMNSRS-Relatedness contains 587 pairs rated by 4 different medical residents. Each dataset can also be divided into 6 semantic categories: DISORDER-DISORDER, SYMPTOM-SYMPTOM, DISORDER-DRUG, DISORDER-SYMPTOM, DRUG-DRUG, and SYMPTOM-DRUG pairs.

In Pakhomov et al.’s evaluation, they modified the UMNSRS dataset to retain only those medical terms that appear in all of the three corpora that they used (Clinical Notes, PubMed Central articles, and Wikipedia). This reduced the number of pairs from 566 to 449 pairs in UMNSRS-Similarity, and from 588 to 458 pairs in UMNSRS-Relatedness.

In our evaluation, we use both the entire UMNSRS dataset and the modified UMNSRS dataset used by Pakhomov et al. For the full dataset, 526 of 566 pairs in UMNSRS-Similarity and 543 of 588 pairs in UMNSRS-Relatedness were found in our pre-processed corpus. For the modified dataset, this corpus contains 418 of 449 pairs for UMNSRS-Similarity and 427 of 458 pairs for UMNSRS-Relatedness.

Semantic Lexicon from UMLS

The Unified Medical Language System is a repository of biomedical vocabularies developed by the US National Library of Medicine. It contains three components: the Metathesaurus; a Semantic Network, and the Specialist Lexicon (lexical information and tools for natural language processing). The Metathesaurus forms the base of the UMLS and comprises over 1 million biomedical concepts. It is organized by concept, and each concept has specific attributes defining its meaning and its links to corresponding concept names in the various source vocabularies [30]. In this work, we only used the UMLS Metathesaurus’ “related concepts” file. This file contains all pair-wise relationships between

concepts (or “atoms”) known to the Metathesaurus. Table 1 displays different relationships and their descriptions.

Table 1– Categories of relationships and their descriptions

Relationship	Description
AQ	has allowed qualifier <i>e.g. Myopathy AQ prevention & control</i>
CHD	has child relationship <i>e.g. Anemia CHD Mild anemia</i>
PAR	has parent relationship <i>e.g. Asthma PAR Bronchial Disease</i>
QB	can be qualified by
RB	has a broader relationship with <i>e.g. Angina RB Pain</i>
RL	the relationship is similar or "alike."
RN	has a narrower relationship with <i>e.g. Hernias RN Hernia, Paraesophageal</i>
RO	has other relationship <i>e.g. Ciprofloxacin RO Cipro 250 MG Tablet</i>
RQ	related and possibly synonymous. <i>e.g. Asthma RQ Wheezing</i>
RU	Related, unspecified
SIB	has sibling relationship <i>e.g. Acne SIB Skin Cancer</i>
SY	the source asserted synonymy <i>e.g. Diarrhea SY Dysentery</i>
XR	not related, no mapping

For each concept in the evaluation dataset, we collected all related concepts within a one-step relationship from this related concepts file. For example, if A is our target concept and we have relationships A CHD B and B CHD C, only B will be considered as a semantic lexicon candidate for A.

Concept-Based Word Embedding Model

To prepare the background corpus for the word embedding model, we downloaded all of the citations (titles and abstracts) in PubMed published before 2016. We then ran SemRep [31], which uses MetaMap [32] for concept extraction and normalization, on each citation’s title and abstract to obtain a sequence of concept unique identifiers (CUI). In other words, following De Vine et al. [23], each sentence in this corpus is replaced by a sequence of CUIs, indicating the order in which concepts were encountered in the text.

To train this word embedding model, we used the *word2vec* implementation in Gensim, a Python package [33] to generate a ‘concept embedding’ for each CUI in our pre-processed corpus. We followed [27] in using the continuous bag-of-words (CBOW) model for word embedding training. The window size was set to 20, and the dimensionality of feature vectors was set at 200. We ignored all CUIs with a total frequency lower than 5.

Retrofitting Word Vector to Semantic Lexicons

Vector space word representations are a critical component of many modern natural language processing systems. Currently it is common practice to represent words using corpus-derived dense high-dimensional vectors. However, this fails to take into account relational structures that have been explicitly encoded into semantic lexicons. Retrofitting is a simple and effective method to improve word vectors using word relationship knowledge encoded in semantic lexicons. It is used as a post-processing step to improve vector quality [28].

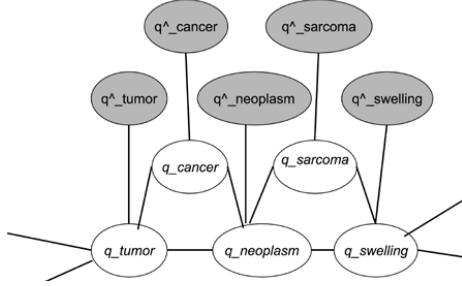


Figure 1– Word graph with edges between related words, observed (gray node), inferred (white node)

Figure 1 shows a small word graph example with edges connecting semantically related words. The words, *cancer*, *tumor*, *neoplasm*, *sarcoma*, and *swelling*, are similar words to each other, as defined in a lexical knowledge resource. Grey nodes represent observed word vectors built from the corpus. White nodes represent inferred word vectors, waiting to be retrofitted. The edge between each pair of white nodes means they represent related words (according to some knowledge source). The inferred word vector (e.g., q_{tumor}) is expected to be close to both its original (pre-retrofitting) estimated word vector (i.e., q^a_{tumor}) and the retrofitted vector of its semantic neighbors (e.g., q_{cancer} and $q_{neoplasm}$). The objective is to minimize the following:

$$\psi(Q) = \sum_{i=1}^n [\alpha_i \|q_i - q_i^a\|^2 + \sum_{(i,j) \in E} \beta_{ij} \|q_i - q_j\|^2]$$

where α and β are hyperparameters that control the relative strengths of corpus- and lexically-derived associations, Q represents the retrofitted vectors, and $(i,j) \in E$ means there is an edge between node q_i and q_j . ψ is convex in Q . An efficient iterative updating method is used to solve the convex objective. First, retrofitted vectors in Q are initialized to be equal to the empirically estimated vectors. The next step is to take the first derivative of ψ with respect to the q_i vector and use the following to update it online.

$$q_i = \frac{\sum_{j:(i,j) \in E} \beta_{ij} q_j + \alpha_i q_i^a}{\sum_{j:(i,j) \in E} \beta_{ij} + \alpha_i}$$

In practice, it takes approximately 10 iterations to converge to the difference in Euclidean distance of adjacent nodes of less than 0.01. We used the authors' implementation of this algorithm [28].

Evaluation Measures

In the evaluation, we tested different semantic lexicons (based on the categories of relationships described in Table 1) with the 'retrofitting' method to improve the vector quality of each concept. For each term pair in the test dataset, we extracted concept vectors and computed the cosine similarity between them using the following equation:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^N A_i B_i}{\sqrt{\sum_{i=1}^N A_i^2} \sqrt{\sum_{i=1}^N B_i^2}}$$

Where A_i and B_i are components of vector A and B respectively, and N is the length of vector. The cosine scores computed for each pair in the test dataset were then compared to the mean of the human similarity and relatedness judgments for each pair, using Spearman rank correlation. We also tested our method on different subsets of the UMNSRS dataset consisting of pairs of different semantic types. The baselines

we used for comparison are the results reported by Pakhomov *et al.* in 2016 [27].

Results

Comparisons with different lexicons

The results of these experiments are shown in Table 2, which shows results after retrofitting for all relationship types in Table 1.. Given differences in vocabulary across corpora, we cannot compare the identical set of pairs used by Pakhomov *et al.* Nonetheless, our CUI-based vector representations based without retrofitting ("No Retrofitting") perform slightly better than the results reported by Pakhomov *et al.* on both full and modified UMNSRS sets. Retrofitting with RO relationships results in best performance for semantic similarity, with correlations of 0.683 and 0.673 for the full and modified datasets, respectively. For UMNSRS-Relatedness, using RQ relationships achieves best performance with correlations of 0.609 and 0.621 for the full and modified dataset, respectively.

Pakhomov <i>et al.</i>	0.62 (n=449)		0.58 (n=458)	
UMNSRS	Similarity		Relatedness	
Test Subsets	Full (n=526)	Modified (n=418)	Full (n=543)	Modified (n=427)
No Retro-fitting	0.639	0.628	0.585	0.594
AQ	0.574	0.552	0.527	0.525
SIB	0.601	0.585	0.530	0.535
PAR	0.632	0.618	0.561	0.562
RB	0.636	0.624	0.586	0.593
RL	0.639	0.628	0.585	0.594
RU	0.639	0.628	0.585	0.594
QB	0.639	0.628	0.585	0.594
XR	0.639	0.628	0.585	0.594
CHD	0.642	0.632	0.588	0.595
SY	0.654	0.644	0.599	0.610
RQ	0.657	0.655	0.609	0.621
RN	0.664	0.656	0.600	0.608
RO	0.683	0.673	0.604	0.613

Table 2-Comparison of Spearman rank correlations between human raters and our method using different lexicons

Only lexical information concerning *CHD*, *SY*, *RQ*, *RN*, and *RO* relationships improved the performance of concept vector representations. Table 3 presents the performance of our retrofitting method using different combinations of productive relationships on the test dataset. Combining *RN* and *RO* relationships resulted in the best performance of 0.689 and 0.681 for the full and modified UMNSRS-Similarity datasets.

For UMNSRS-Relatedness, lexicons with all five productive relationships attained the highest correlations of 0.624 and 0.635 for the full and modified datasets respectively. Furthermore, any lexicons including *RO* relationship have similar performance for UMNSRS-Similarity and any lexicons including *RQ* obtain similar correlation scores for UMNSRS-Relatedness.

Table 3-Comparison of Spearman rank correlations between human raters and our method using lexicons combinations

Lexicons Combinations	0.62 (n=449)		0.58 (n=458)	
	UMNSRS-Similarity		UMNSRS-Relatedness	
	Full (n=526)	Modified (n=418)	Full (n=543)	Modified (n=427)
No Lexicons	0.639	0.628	0.585	0.593
CHD+SY	0.651	0.643	0.596	0.605
RQ+RO	0.686	0.679	0.616	0.627
RN+RQ	0.667	0.662	0.607	0.617
RN+RO	0.689	0.681	0.619	0.630
RN+RO+RQ	0.687	0.681	0.622	0.634
SY+RN+RO+RQ	0.686	0.680	0.623	0.634
CHD+SY+RN+RO+RQ	0.686	0.680	0.624	0.635

Comparison across pairs of different semantic types

From Table 3, we can see that lexicons containing *RN+RO* and *CHD+SY+RN+RO+RQ* achieved the best performances for UMNSRS-Similarity and UMNSRS-Relatedness respectively. Hence, we just used these two lexicons in the comparison of Spearman rank correlations between human raters and our method in different subsets of pairs grouped by semantic types. Table 4 and Table 5 present performances of comparisons for UMNSRS-Similarity and UMNSRS-Relatedness. As shown in Table 4, the lexicon from *RN* and *RO* relationships achieves the best correlation performance in 4 of 6 groups and lexicon from *CHD*, *SY*, *RN*, *RO*, and *RQ* relationships obtain the highest correlation score in symptom-symptom pairs. However, Pakhomov *et al.* retain the best performance for disorder-disorder (Di-Di) pairs, using PMC.

Table 4-Comparison of Spearman rank correlations between human raters estimates of similarity and our method in different subsets of pairs grouped by semantic types (Di-disorder, S-symptom, Dr-drug)

UMNSRS-Similarity	Pakhomov et al.	RN+RO		CHD+SY+RN+RO+RQ	
		Mod	Full	Mod	Full
All Pairs	0.62	0.681	0.689	0.680	0.686
Di-Di	0.74	0.715	0.72	0.723	0.726
S-S	0.56	0.625	0.668	0.635	0.670
Dr-Dr	0.77	0.841	0.749	0.840	0.748
Di-S	0.49	0.703	0.720	0.699	0.717
Di-Dr	0.69	0.686	0.710	0.682	0.708
S-Dr	0.51	0.484	0.552	0.476	0.546

As shown in Table 5, the lexicon containing *CHD*, *SY*, *RN*, *RO*, and *RQ* relationships resulted in the highest correlation with human raters in 4 of 6 groups for UMNSRS-Relatedness dataset. Pakhomov *et al.* retained the best performance in disorder-drug (Di-Dr) and symptom-drug (S-Dr), achieved using embeddings trained on clinical notes [27].

Discussion

In this study, we used a method for retrofitting of word embeddings to improve semantic similarity and relatedness measures by incorporating structural information from the

UMLS. We evaluated our approach on both the full UMNSRS dataset and the modified subset used in [27]. Vector representations trained on sequences of CUIs (without retrofitting) resulted in comparable performance (with slight improvements) to those based on sequences of terms. After applying retrofitting on CUI vector represents using selected UMLS relationship types, we see clear improvements on both the full and modified dataset, compared to the CUI vectors without retrofitting.

Table 5- Comparison of Spearman rank correlations between human raters estimates of relatedness and our method in different subsets of pairs grouped by semantic types (Di-disorder, S-symptom, Dr-drug)

UMNSRS-Relatedness	Pakhomov et al.	RN+RO		CHD+SY+RN+RO+RQ	
		Mod	Full	Mod	Full
All Pairs	0.58	0.630	0.619	0.635	0.624
Di-Di	0.59	0.589	0.628	0.593	0.629
S-S	0.64	0.692	0.706	0.724	0.730
Dr-Dr	0.73	0.734	0.571	0.736	0.572
Di-S	0.42	0.562	0.594	0.569	0.603
Di-Dr	0.63	0.564	0.607	0.565	0.611
S-Dr	0.59	0.479	0.519	0.482	0.523

In comparison with the best results previously reported on the UMNSRS benchmark ([27] - 0.62 for similarity and 0.58 for relatedness), we obtain better correlation with human raters on both similarity and relatedness (0.689 for similarity and 0.624 for relatedness on the full UMNSRS dataset and 0.681 for similarity and 0.635 for relatedness on the modified version). However, as our results concern a subset of the modified set only, further evaluation on matching sets is required to show this conclusively. Our codes and word embeddings are available at (<https://github.com/Sssssstanley/Retrofitting-Concept-Vector-Representations-of-Medical-Concepts>).

However, our results also show that external linkage information should be carefully chosen. For example, using *AQ*, *SIB*, *PAR*, and *RB* relationships resulted in worse correlation with human judgment than the original concept vectors (without retrofitting). This suggests that these relationship types are too permissive to align with human evaluation. Incorporating other relationships, such as *RB*, *RL*, *RU*, and *XR*, had no effect on the results. The reason for this is that no CUIs connected to CUIs in the evaluation set using these relationships. *CHD*, *SY*, *RQ*, *RN*, and *RO* clearly have positive effects on the quality of the vector representations. *RO* has the largest positive effect on the Similarity dataset, and *RQ* improves the vector presentation the most on the Relatedness dataset. The description of *RO* is 'has a relationship other than synonymous, narrower, or broader.' For example, *Ciprofloxacin* and *Cipro 250 MG Oral Tablet* are linked by *RO*. These are the same drug with different dosages, so retrofitting would enhance similarity between vectors for concepts representing the same drug. The description of *RQ* is 'related and possibly synonymous'. "Relatedness" is a general notion that encompasses similarity, and maps well to this relationship type. Hence, it seems reasonable that incorporating this relation would achieve the best correlation with human raters on UMNSRS-Relatedness dataset.

As noted in [27] the correlations in the 0.5~0.6 range reported for the UMNSRS benchmark are in the same range as the intra-class correlation coefficients used to measure agreement between human annotators for this set, and so may constitute the ceiling for performance that can be measured using this benchmark. However, our results are clearly over this range. What we reported are correlations with the mean rating, which

may be more readily approximated than the ratings of a single rater. In the future, we will conduct further analysis on interpreting our results in relation to the inter-rater agreement intra-class correlations for different categories of term pairs.

Conclusions

In this paper, we introduced a hybrid method for generating semantic vector representations of UMLS concepts, by leveraging both distributional statistics and linkage information from an ontology or taxonomy (such as the UMLS). This method achieved better performance on the UMNSRS benchmark than neural word embedding alone, with the best results reported for this evaluation to date. Any application using concept vector representations could potentially benefit from the additional structural information encoding using this retrofitting approach. In the future, we will continue to evaluate the utility of retrofitting method for downstream tasks (such as word-sense disambiguation and information retrieval).

Acknowledgements

This work was supported by the UTHHealth Innovation for Cancer Prevention Research Training Program Predoctoral Fellowship (Cancer Prevention and Research Institute of Texas (CPRIT) grant # RP160015), and supported in part by National Library of Medicine R01LM011563. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CPRIT or NLM.

References

- [1] R. Rada, H. Mili, E. Bicknell, and M. Blettner, Development and application of a metric on semantic nets, *IEEE transactions on systems, man, and cybernetics* **19** (1989), 17-30.
- [2] T. Cohen and D. Widdows, Empirical distributional semantics: methods and biomedical applications, *Journal of biomedical informatics* **42** (2009), 390-405.
- [3] Y. Lin, W. Li, K. Chen, and Y. Liu, A document clustering and ranking system for exploring MEDLINE citations, *Journal of the American Medical Informatics Association* **14** (2007), 651-661.
- [4] P. Agarwal and D.B. Searls, Can literature analysis identify innovation drivers in drug discovery?, *Nature Reviews Drug Discovery* **8** (2009), 865-878.
- [5] F. Doshi-Velez, B. Wallace, and R. Adams, Graph-sparse LDA: a topic model with structured sparsity, *arXiv preprint arXiv:1410.4510* (2014).
- [6] O. Bodenreider and A. Burgun, Aligning knowledge sources in the UMLS: methods, quantitative results, and applications, *Studies in health technology and informatics* **107** (2004), 327.
- [7] T. Pedersen, S.V. Pakhomov, S. Patwardhan, and C.G. Chute, Measures of semantic similarity and relatedness in the biomedical domain, *Journal of biomedical informatics* **40** (2007), 288-299.
- [8] S.V. Pakhomov, T. Pedersen, B. McInnes, G.B. Melton, A. Ruggieri, and C.G. Chute, Towards a framework for developing semantic relatedness reference standards, *Journal of biomedical informatics* **44** (2011), 251-265.
- [9] J.E. Caviedes and J.J. Cimino, Towards the development of a conceptual distance metric for the UMLS, *Journal of biomedical informatics* **37** (2004), 77-85.
- [10] T. Pedersen, S. Patwardhan, and J. Michelizzi, WordNet:: Similarity: measuring the relatedness of concepts, in: *Demonstration papers at HLT-NAACL 2004*, Association for Computational Linguistics, 2004, pp. 38-41.
- [11] M. Ciaramita, A. Gangemi, and E. Ratsch, Unsupervised learning of semantic relations for molecular biology ontologies, *Ontology Learning and Population: Bridging the Gap between Text and Knowledge* (2008), 91-107.
- [12] P.D. Turney, Measuring semantic similarity by latent relational analysis, *arXiv preprint cs/0508053* (2005).
- [13] V.N. Garla and C. Brandt, Semantic similarity in the biomedical domain: an evaluation across knowledge sources, *BMC bioinformatics* **13** (2012), 261.
- [14] C. Leacock and M. Chodorow, Combining local context and WordNet similarity for word sense identification, *WordNet: An electronic lexical database* **49** (1998), 265-283.
- [15] Z. Wu and M. Palmer, Verbs semantics and lexical selection, in: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, Association for Computational Linguistics, 1994, pp. 133-138.
- [16] D. Lin, An information-theoretic definition of similarity, in: *ICML*, Citeseer, 1998.
- [17] S. Patwardhan and T. Pedersen, Using WordNet-based context vectors to estimate the semantic relatedness of concepts, in: *Proceedings of the eacl 2006 workshop making sense of sense-bringing computational linguistics and psycholinguistics together*, 2006.
- [18] S. Pakhomov, B. McInnes, T. Adam, Y. Liu, T. Pedersen, and G.B. Melton, Semantic similarity and relatedness between clinical terms: an experimental study, in: *AMIA annual symposium proceedings*, American Medical Informatics Association, 2010, p. 572.
- [19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, Efficient estimation of word representations in vector space, *arXiv preprint arXiv:1301.3781* (2013).
- [20] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, in: *Advances in neural information processing systems*, 2013, pp. 3111-3119.
- [21] M. Baroni, G. Dinu, and G. Kruszewski, Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors, in.
- [22] O. Levy, Y. Goldberg, and I. Dagan, Improving distributional similarity with lessons learned from word embeddings, *Transactions of the Association for Computational Linguistics* **3** (2015), 211-225.
- [23] L. De Vine, G. Zucco, B. Koopman, L. Sitbon, and P. Bruza, Medical semantic similarity with a neural language model, in: *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, ACM, 2014, pp. 1819-1822.
- [24] P. Kanerva and A. Holst, Random indexing of text samples for latent semantic analysis, in, Citeseer, 2000.
- [25] T.K. Landauer and S.T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge, *Psychological review* **104** (1997), 211.
- [26] A. Sajadi, Graph-Based Domain-Specific Semantic Relatedness from Wikipedia, in: *Canadian Conference on Artificial Intelligence*, Springer, 2014, pp. 381-386.
- [27] S.V. Pakhomov, G. Finley, R. McEwan, Y. Wang, and G.B. Melton, Corpus domain effects on distributional semantic modeling of medical terms, *Bioinformatics* **32** (2016), 3635-3644.
- [28] M. Faruqi, J. Dodge, S.K. Jauhar, C. Dyer, E. Hovy, and N.A. Smith, Retrofitting word vectors to semantic lexicons, *arXiv preprint arXiv:1411.4166* (2014).
- [29] Z. Yu, T. Cohen, E.V. Bernstam, T.R.J. MSE, and B.C. Wallace, Retrofitting Word Vectors of MeSH Terms to Improve Semantic Similarity Measures, *EMNLP 2016*, 43.
- [30] O. Bodenreider, The unified medical language system (UMLS): integrating biomedical terminology, *Nucleic acids research* **32** (2004), D267-D270.
- [31] T.C. Rindfleisch and M. Fiszman, The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypemymic propositions in biomedical text, *Journal of biomedical informatics* **36** (2003), 462-477.
- [32] A.R. Aronson, Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program, in: *Proceedings of the AMIA Symposium*, American Medical Informatics Association, 2001, p. 17.
- [33] R. Rehurek and P. Sojka, Software framework for topic modelling with large corpora, in: *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Address for correspondence

Trevor Cohen, MBChB, PhD
7000 Fannin St, Suite 600, Houston, TX, 77030
Email : Trevor.Cohen@uth.tmc.edu
Phone : 713.486.3675 Fax : 713.486.0117

"Hybrid Topics" -- Facilitating the Interpretation of Topics Through the Addition of MeSH Descriptors to Bags of Words

Zhiguo Yu^a, Thang Nguyen^b, Ferdinand Dhombres^c, Todd Johnson^a, Olivier Bodenreider^c

^aThe University of Texas at Biomedical Informatics at Houston, Houston, Texas, USA,

^bDepartment of Computer Science, University of Maryland, College Park, Maryland, USA

^cU.S. National Library of Medicine, National Institute of Health, Bethesda, Maryland, USA

Abstract

Extracting and understanding information, themes and relationships from large collections of documents is an important task for biomedical researchers. Latent Dirichlet Allocation is an unsupervised topic modeling technique using the bag-of-words assumption that has been applied extensively to unveil hidden thematic information within large sets of documents. In this paper, we added MeSH descriptors to the bag-of-words assumption to generate 'hybrid topics', which are mixed vectors of words and descriptors. We evaluated this approach on the quality and interpretability of topics in both a general corpus and a specialized corpus. Our results demonstrated that the coherence of 'hybrid topics' is higher than that of regular bag-of-words topics in the specialized corpus. We also found that the proportion of topics that are not associated with MeSH descriptors is higher in the specialized corpus than in the general corpus.

Keywords:

Medical Subject Headings; Models, Statistical Data; Data Interpretation, Statistical

Introduction

Motivation

Knowledge discovery is a fundamental and important activity in biomedical research. Given that unstructured data is increasing exponentially, extracting and understanding information, themes, and relationships from large collections of documents is increasingly important. PubMed currently comprises more than 26 million articles from the biomedical literature and has been widely used to help researchers keep up with the state of the art in their domains and explore other unfamiliar research areas. To help users better retrieve relevant information and manage this tremendous volume of biomedical literature, the US National Library of Medicine (NLM) has developed the Medical Subject Headings (MeSH) controlled vocabulary for indexing MEDLINE articles. MeSH has been used to improve PubMed query results [1-2]. However, users are still often overloaded by the tremendous number of relevant articles returned from their PubMed queries [3]. Hence, biomedical researchers need an efficient and convenient way to discover knowledge from large sets of documents.

MeSH Indexing vs. Topics

Latent Dirichlet Allocation (LDA) [4], is a popular topic modeling method that aims to extract the semantic themes (topics)

automatically from a corpus of documents. These topics describe the thematic composition of documents and can thus capture the semantic similarity between them. In contrast, MeSH descriptors are manually created and maintained by domain experts to cover all important themes. Topics extracted from a subset of documents are subset-specific [5]. Thus, they may identify corpus-specific themes that may not be covered in MeSH. Such themes may uncover a specific set of concepts for a particular domain or sub-domain.

Related Work

Considerable research has examined the application of topic models to MeSH descriptors. Labeled LDA (labeled LDA) [6] is a supervised topic model developed to uncover latent topics that correlate with user tags in labeled corpora. In other words, each tag will be represented as a topic. Zhu et al. have used labeled LDA in an attempt to automatically assign MeSH descriptors to new publications (not yet indexed with MeSH descriptors) [7]. Elsewhere, Newman et al. presented a resampled author model that combines both general LDA and the author-topic model (in this case MeSH descriptors were treated as the 'authors'). The resampled author model provided an alternative and complementary view of the relationships between MeSH descriptors [8]. All these investigations used topic models to interpret MeSH descriptors. However, they cannot be used to identify themes that are not covered in MeSH. In 2014, a graph-sparse LDA model [9] was developed to generate more interpretable topics by leveraging relationships expressed by controlled vocabulary structure (e.g. MeSH). In this model, a few concept-words from the controlled vocabulary can be identified to represent generated topics. Though MeSH was shown to work well for summarizing biomedical articles, this model cannot identify themes that may not exist in the MeSH vocabulary.

Specific Contribution

This work introduces an alternative LDA approach by migrating its original bag-of-words assumption to a 'bag-of-MeSH&words' approach. By enriching each document with its indexed MeSH descriptors, 'hybrid topics' (mixed vectors of words and MeSH descriptors) can be generated by LDA.

Objectives

We investigated whether the addition of labels (e.g., MeSH descriptors) to bags-of-words can improve the quality and facilitate the interpretation of LDA-generated topics. More specifically, to assess the quality and interpretability of topics, we tested two hypotheses using one large general biomedical corpus and one smaller specialized biomedical corpus.

1. The coherence (used as a surrogate for quality) of 'hybrid topics' is expected to be higher than that of regular bag-of-words topics.
2. The proportion of topics that are not associated with some MeSH descriptor, which reflects limited interpretability, is expected to be higher in a specialized corpus than in a general corpus.

Background

Medical Subject Headings (MeSH)

The MeSH controlled vocabulary was developed by NLM to help manage, index, and search MEDLINE articles. There were 27,883 descriptors in the 2016 MeSH, with over 87,000 entry terms that assist in finding the most appropriate MeSH descriptor (for example, 'Vitamin C' is an entry term to 'Ascorbic Acid'). In the 2016 MeSH, 82 qualifiers could be attached to MeSH descriptors to describe a particular aspect of a concept, such as 'adverse effects', 'diagnosis', etc. Each year, the MeSH specialists revise and update the MeSH vocabulary to cover emerging research areas and improve indexing consistency and efficiency. MeSH specialists are responsible for areas of the health sciences in which they have knowledge and expertise. MEDLINE indexers make suggestions for new descriptors to MeSH specialists during their indexing processes. Besides, MeSH specialists also collect new terms as they appear in the scientific literature or emerging areas of research. After defining these terms within the context of the existing vocabulary, MeSH specialists recommend their addition(s) to MeSH. During each MEDLINE year-end processing (YEP) activities, changes made to MeSH are applied to MEDLINE (retrospectively) for conformance with the current version of MeSH.

Latent Dirichlet Allocation (LDA)

LDA is a generative model that assumes that each document is generated from a mixture of topics and that each topic corresponds to a distribution over all words in the corpus. Informally, the 'generative story' for LDA is as follows. First, a document is generated by drawing a mixture of topics that the document is about. To generate each word in this document, one draws a topic from this distribution and subsequently selects a word from the distribution over the vocabulary of the whole corpus corresponding to this topic. The LDA algorithm uses this generative model to uncover the latent topics contained within a given corpus. Specifically, it estimates the parameters that define document topic mixtures and the conditional probabilities of each word given each topic. Parameter estimation is usually done via sampling approaches.

The number of topics produced by LDA must be prespecified. Determining the 'right' number of topics for different datasets remains a challenge. When the number of topics increases, redundant and nonsense topics may be generated. Running LDA with a small number of topics will generate more general themes. In this paper, we used a topic coherence measure to determine the optimal number of topics for our dataset [10]. Details are described in the next section.

Quality of Topics

Recently, O'Callaghan et al. [10] reviewed a number of topic coherence studies using various corpora and proposed a general measure based on distributional semantics, *TC-W2V*. This measure evaluates the relatedness of a set of top terms describing a topic, based on the similarity of their representations in a *word2vec* distributional semantic space. Specifically, the coherence of a topic n represented by its top t ranked

terms is given by the mean pairwise cosine similarity between all relevant term vectors in the *word2vec* space:

$$coh(t_n) = \frac{1}{\binom{t}{2}} \sum_{j=2}^t \sum_{i=1}^{j-1} \cos(wv_i, wv_j)$$

An overall score for a topic model M consisting of k topics is calculated by averaging the individual topic coherence scores:

$$coh(M) = \frac{1}{k} \sum_{n=1}^k coh(t_n)$$

In this investigation, we use topic coherence to help determine the optimal number and quality of topics.

Relations between MeSH and Topics

Topics are generated based on the bag-of-words assumption, which ignores word order. Each topic is represented as a list of ranked words, which provides the user with a sense of what the topic is about. Each document is displayed as a list of weighted topics, which represents different aspects of the document. Since the tokens within each topic are ranked according to the conditional probabilities $P(w|t)$ learned when training the model, where w is a word and t is a topic, the top few words of each topic provides insight into the subject of the topic. However, the interpretation of the topics (i.e., lists of words) is left as an exercise for the user.

As mentioned earlier, MeSH was developed to cover all important themes and each article in MEDLINE is indexed with a few relevant MeSH descriptors assigned by the MEDLINE indexing staff for retrieval purposes. To capture the potential relationships between MeSH and topics, we simply added the MeSH descriptors assigned to each article to the bag-of-words for a given document, creating a modified 'bag-of-MeSH&word'. Under this 'bag-of-MeSH&word' assumption, 'hybrid topics' are generated and each topic is represented as a list of tokens (i.e., a mixed list of ranked words and MeSH descriptors). The presence of MeSH descriptors among the top tokens for a given topic is expected to facilitate the interpretation of topics. More specifically, if a MeSH descriptor appears among the top m (for some m) tokens of a topic, we will assume this MeSH descriptor is highly associated with this topic. We consider three types of association patterns:

1. One topic has no MeSH descriptor in its top m tokens (1-0 mapping);
2. One topic has a single MeSH descriptor in its top m tokens (1-1 mapping);
3. One topic has multiple MeSH descriptors in the top m tokens (1-many mapping).

Examples of topics for each association pattern are shown in Table 1 along with the top 10 tokens for each topic.

Table 1—Generated 'bag-of-MeSH&word' topics (*MeSH descriptors)

Topic 1	Topic 2	Topic 3
model	*brain	motor
predict	cortex	visual
value	region	*movement
prediction	functional	*face
analysis	cortical	right
predictive	activity	response
regression	neural	*hand
datum	network	processing
estimate	change	object
predictor	area	stimuli
1-0 mapping	1-1 mapping	1-many mapping

Methods

Data Preparation

One large general corpus and one small specialized corpus were used in this investigation. The general corpus consisted of 200k articles randomly selected from all PubMed articles published in 2013. The specialized corpus consisted of 2,472 articles from the journal *Prenatal Diagnosis*, which focuses on fetal medicine.

General Corpus

There are roughly 1.2 million articles in PubMed for the year 2013. We randomly selected 200k articles (titles and abstracts) from these. This represents an appropriate amount of data given our computing resources.

To reduce the sparsity of the document-to-words distribution, we performed Part of Speech tagging on the dataset and merged several categories, including *NN* and *NNS* (e.g., patient and patients); *VB*, *VBD*, *VBG*, and *VBN* (e.g., eat, ate, eaten, and eating); and *JJ*, *JJR*, and *JJS* (e.g., good, better, and best).

We also removed PubMed stop-words and infrequent words (with a frequency lower than 50). A total of 21,922 unique words remained. Similarly, for MeSH descriptors, we treated specific frequently used descriptors known as check tags (e.g., human, male, female, etc.) as stop words, and ignored infrequent descriptors (with a frequency lower than 5). A total of 13,853 MeSH descriptors remained.

Specialized Corpus

We applied similar preprocessing to the specialized corpus, but with different cutoff values due to its smaller size. After setting a cutoff frequency of five for words, we obtained 3,623 unique words. With a cutoff frequency of one for MeSH descriptors, we obtained 919 MeSH descriptors.

Experiment #1

We investigated whether the addition of MeSH descriptors to bags-of-words increased the quality of topics. As a surrogate for the quality of topics, we used topic coherence [11].

In practice, to determine whether our 'hybrid topics' approach (i.e., 'bag-of-MeSH&words') outperformed the original LDA bag-of-words approach (baseline), we generated LDA models under both these assumptions for a various number of topics on the two datasets.

For the general corpus, the number of topics tested 50-600. For the specific corpus, we tested 4-100. For each number of topics, we calculated topic coherence for both the baseline and the 'hybrid topics' approaches.

More specifically, for the large general corpus, we used the indexed PubMed articles (titles and abstracts) published in 2013 as our background corpus when building the *word2vector* space for the original LDA with the bag-of-words assumption. To build the *word2vector* space containing both MeSH descriptors and words, we simply appended the MeSH descriptors for an article to the end of the document. In this way, we could get a mixed *word2vector* space of MeSH descriptors and words. In our experiment, we tested two different positions of MeSH descriptors in the citation (front and end) and obtained similar topic coherence results. Following [10], we used the same *word2vec* setting and the number of top terms per topic ($t=10$).

For the small specialized corpus, we used the full-text of these articles as the background corpus when building the *word2vec* space. To build the mixed *word2vector* space of MeSH descriptors and words for this background corpus, we added MeSH descriptors to the end of each full article. In the *word2vec* setting for this dataset, we set vector size to 200, cutoff frequency to three, and window size to 20.

To compare the topic coherence measures obtained within each corpus at different numbers of topics for the baseline and the 'hybrid topics', we used a paired t-test.

Experiment #2

To assess whether the proportion of 'hybrid topics' that were not associated with some MeSH descriptor, which reflects limited interpretability, were higher in a specialized corpus than in a general corpus, we first had to determine the optimal number of topics in each corpus.

Choosing the number of topics k is a key parameter selection decision in topic modeling. Too few topics will produce results that are overly broad, while too many will lead to many small, highly similar topics. One general strategy proposed in the literature is to compare the topic coherence of topic models with different values of k . An appropriate value for k can then be identified by examining a plot of the mean *TC-W2V* coherence scores for a fixed range and selecting a value corresponding to the maximum coherence score. Since we only expected the MeSH descriptors to help interpret topics rather than introduce new topics, we just used LDA's original bag-of-words assumption to determine the optimal number of topics for each test corpus.

Having determined the optimal number of topics for each corpus, we examined the 'hybrid topics' obtained for this number of topics and counted which ones were not associated with MeSH descriptors (i.e., which ones did not contain at least one MeSH descriptor among their top-20 tokens).

We used the chi-square statistic to compare the distribution of topics of two patterns between the 'hybrid topics' and the baseline.

Results

Experiment #1

Figures 1 and 2 display the difference in topic coherence between our 'bag-of-MeSH&words' assumption ('hybrid topics') and LDA's original bag-of-words assumption (baseline) for the general and specialized corpora respectively.

For the general corpus, we computed topic coherence for 10 different numbers of topics for both the baseline and our 'hybrid topics'. As shown in Figure 1, topic coherence scores were very close between the baseline and 'hybrid topics'. The coherence was slightly better with 'hybrid topics' after 100 topics, but slightly lower for 50 and 100 topics.

For the specialized corpus, however, we saw a clear improvement on the coherence of topics in favor of 'hybrid topics' compared to the baseline. As shown in Figure 2, topic coherence scores were systematically higher for 'hybrid topics' across all numbers of topics.

Though 'hybrid topics' are higher the baseline after 100 topics on the general corpus, the paired t-test was not significant ($p=0.1624$). We could not properly assess the difference between the two approaches on this general corpus. With the specialized corpus, however, the paired t-test was highly sig-

nificant ($p=6.8e-25$), demonstrating that the quality of the 'hybrid topics' was better than that of the baseline topics.

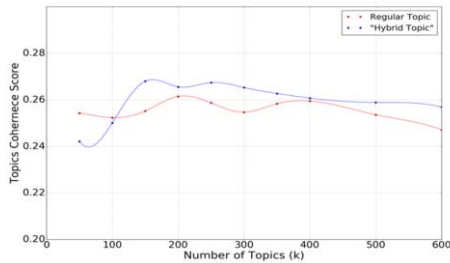


Figure 1- Comparison of mean TC-W2V topic coherence scores for different numbers of topics k , generated from the general corpus

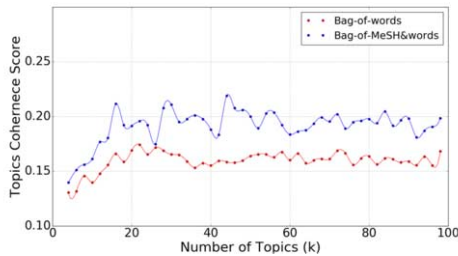


Figure 2- Comparison of mean TC-W2V topic coherence scores for different numbers of topics k , generated from the specialized corpus

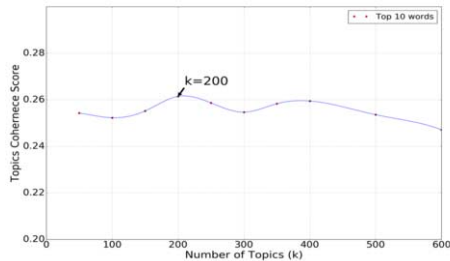


Figure 3- Plot of mean TC-W2V topic coherence scores for different numbers of topics k , generated from the general corpus

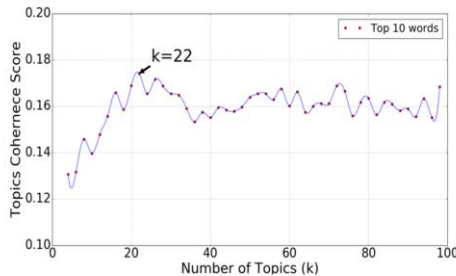


Figure 4- Plot of mean TC-W2V topic coherence scores for different numbers of topics k , generated from the specialized corpus

Experiment #2

Optimal Number of Topics

For the large general corpus, we generated LDA models containing $k \in [50, 600]$ topics and selected the value of k that maximized the mean TC-W2V coherence. As shown in Figure 3, $k=200$ is the first maximum and was therefore selected as the optimal number of topics for this dataset.

For the small specialized corpus, we generated LDA models containing $k \in [4, 100]$ topics. As shown in Figure 4, $k=22$ was the first maximum and was therefore selected as the optimal number of topics for this dataset.

Proportion of Topics Not Associated with a Mesh Descriptor

Table 2 displays the number of different patterns of association between topics and MeSH descriptors observed in the general and specialized corpus for their respective optimal number of topics.

As shown in Table 2, the proportion of topics not associated with a MeSH descriptor was higher in the specialized corpus (41%) than in the general corpus (3%). The chi-square statistics was 57.36 ($p=3.502e-13$), which suggests that the corpora had significantly different distributions.

Table 2– # of topics with 0, 1, n MeSH descriptors ($n > 1$)

Data Set	Optimal K	# of Topic with 0 MD	# of Topic with 1 MD	# of Topic with n MD
General Corpus	200	6(3%)	16(8%)	178(89%)
Spec. Corpus	22	9(41%)	6(27%)	7(32%)

Discussion

Findings and Significance

This investigation demonstrates that the addition of MeSH descriptors to the traditional bag-of-words approach to creating topic models ('hybrid topics') can improve the quality of the topics and facilitate their interpretation, but the impact is different on a general corpus and on a specialized corpus. The quality of the 'hybrid topics', assessed by their coherence, was better than that of the baseline topics in the specialized corpus, but it did not seem to be the case in the general corpus.

MeSH terms are created and maintained by MeSH specialists to cover all general themes in biomedicine. However, topics extracted from a subset of documents are often specific to these documents. For the general corpus, most of the topics captured by LDA were general themes. Hence, this addition of MeSH descriptors to the bag-of-words approach did not contribute too much to the topic quality. This could be the reason that we did not see a significant improvement of the topic coherence score between regular and 'hybrid topics' in the general corpus. In contrast, for the specialized corpus, adding MeSH descriptors can provide additional information for LDA to better differentiate between general and specific themes and to improve topics quality.

In terms of interpretability, however, the general corpus benefited from 'hybrid topics' more than the specialized corpus did, because over 40% (9/22) of the 'hybrid topics' remained unlabeled (i.e., not associated with any MeSH descriptors) in the specialized corpus, compared to 3% (6/200) in the general corpus.

Applications to Corpus Exploration

From the general corpus, we saw that only 6/200 topics (3%) contained no MeSH descriptors in their top 20 terms. For the specialized corpus, 9/22 topics (40%) were generated with no MeSH descriptors in their top 20 terms. General themes from the MeSH vocabulary may not be able to cover in detail all aspects of a specialized corpus. In contrast, the topics generated by LDA from a corpus are corpus-specific. It is therefore logical that more topics with no MeSH descriptors are generated from a specialized corpus than a general corpus. Hence LDA will be more useful for a specialized corpus on the task of exploring concepts that may not be covered by MeSH.

From the general corpus, we also saw that 178/194 topics were associated with MeSH descriptors (92%) generated with multiple MeSH descriptors. MeSH descriptors were characterized in 16 top-level categories, such as category A for anatomic terms, category B for organisms, C for diseases, etc. Of these 178 topics, 140 (79%) contained MeSH descriptors from different top-level MeSH categories. These topics were most likely interdisciplinary topics. For the specialized corpus, 7/13 topics associated with MeSH descriptors (54%) were generated with multiple MeSH descriptors. Topics associated with multiple MeSH descriptors from different top-level MeSH categories could be used to explore the intersection of multiple domains. LDA clearly offers an advantage for discovering interdisciplinary topics.

Limitations and Future Work

One limitation of this work is that we ignored the MeSH qualifiers and only considered the MeSH descriptors when constructing our 'hybrid topics'. In the future, we will include the qualifiers to our 'hybrid topics' to test whether it improves the interpretation of topic models. We are also planning to run LDA with a larger number of topics.

Conclusion

In this paper, we introduced an alternative LDA model by adding labels (here, MeSH descriptors) to the bag-of-words assumption. With this setting, 'hybrid topics' can be generated to reveal relationships between topics and labels. In our evaluation, these 'hybrid topics' resulted in higher topic coherence scores compared the original LDA, but only on the specialized biomedical corpus. For the general corpus, we did not see a significant difference on topic quality between our 'hybrid topics' and baseline topics. From our results, we can also conclude that LDA is more useful in the specialized corpus to explore concepts that may not be covered by the MeSH vocabulary and where topic models can capture aggregate concepts from different domains.

Topic models have a strong potential for analyzing the content of large text corpora. However, the deployment of topic models in the real world has been limited. Our goals in the future are to find more practical ways to apply topic models to help people better understand the massive amount of unstructured data available to us.

Acknowledgements

This work was supported by the UTHealth Innovation for Cancer Prevention Research Training Program Predoctoral Fellowship (Cancer Prevention and Research Institute of Texas (CPRIT) grant # RP160015), the Intramural Research Program of the NIH, National Library of Medicine (NLM), and

the NLM Medical informatics training program for graduate and medical students. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CPRIT.

References

- [1] Z. Lu, W. Kim, and W.J. Wilbur, Evaluation of query expansion using MeSH in PubMed, *Information retrieval* **12** (2009), 69-80.
- [2] R.R. Richter and T.M. Austin, Using MeSH (medical subject headings) to enhance PubMed search strategies for evidence-based practice in physical therapy, *Physical therapy* **92** (2012), 124-132.
- [3] R.I. Dogan, G.C. Murray, A. Névél, and Z. Lu, Understanding PubMed® user search behavior through log analysis, *Database* **2009** (2009), bap018.
- [4] D.M. Blei, A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research* **3** (2003), 993-1022.
- [5] Z. Yu, T.R. Johnson, and R. Kavuluru, Phrase based topic modeling for semantic information processing in biomedicine, in: *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, IEEE, 2013, pp. 440-445.
- [6] D. Ramage, D. Hall, R. Nallapati, and C.D. Manning, Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora, in: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, Association for Computational Linguistics, 2009, pp. 248-256.
- [7] D. Zhu, D. Li, B. Carterette, and H. Liu, An Incremental Approach for MEDLINE MeSH Indexing, in: *BioASQ@CLEF*, Citeseer, 2013.
- [8] D. Newman, S. Karimi, and L. Cavedon, Using topic models to interpret MEDLINE's medical subject headings, in: *Australasian Joint Conference on Artificial Intelligence*, Springer, 2009, pp. 270-279.
- [9] F. Doshi-Velez, B. Wallace, and R. Adams, Graph-sparse LDA: a topic model with structured sparsity, *arXiv preprint arXiv:1410.4510* (2014).
- [10] D. O'Callaghan, D. Greene, J. Carthy, and P. Cunningham, An analysis of the coherence of descriptors in topic modeling, *Expert Systems with Applications* **42** (2015), 5645-5657.
- [11] J.H. Lau, D. Newman, and T. Baldwin, Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality, in: *EACL*, 2014, pp. 530-539.

Address for correspondence

Olivier Bodenreider, MD, PhD
8600 Rockville Pike, 38A/ 09S904, Bethesda, MD 20894
Phone Number: (301) 827-4982
E-mail: olivier@nlm.nih.gov

Identifying Biomarkers of Hepatocellular Carcinoma Based on Gene Co-Expression Network from High-Throughput Data

Ying Zhang^{a,b}, Zhiping Liu^b, Jing-song Li^a

^a Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang Province, China

^b Biomedical Engineering, College of Control Science and Engineering, Shandong University, Jinan, Shandong Province, China

Abstract

In this paper, we proposed an approach systematically based on the use of gene co-expression network analyses to identify potential biomarkers for Hepatocellular Carcinoma (HCC). With the analysis of differential gene expression, we first selected candidate genes closely related to HCC from the whole genome on a large scale. By identifying the relationships between each two genes, we built up the gene co-expression network using Cytoscape software. Then the global network was clustered into several sub-modules by Markov Cluster Algorithm (MCL). And, GO-Analysis was carried out for these identified gene modules to further explore the genes obviously associated with the dysfunctions of HCC, and in result we find Hexokinase 2 (HK2) and Krippel-like Factor 4 (KLF4) as potential candidate biomarkers to provide insights into the mechanism of the development of HCC. Finally, we evaluated the disease classification results via an SVM-based machine learning method to verify the accuracy of the classification

Keywords:

Hepatocellular Carcinoma; Cluster Analysis; Machine Learning

Introduction

HCC is an aggressive disease with a high morbidity rate and mortality rate, which is a serious threat to people's lives and health. At present, the treatment of HCC is still limited. More than half of the patients will relapse in the case of surgical resection and will suffer from more and more serious postoperative complications [1]. We can see the overall therapeutic effect of HCC is not optimistic. In addition, two thirds of patients with HCC are in the advanced stage when they are treated, then surgical resection is not the best choice. Therefore, it is particularly important to strengthen the research on the occurrence and development mechanism of HCC to effectively identify the biomarkers for clinical detection of liver cancer, so that we will achieve early detection and treatment of HCC and prolong survival time of patients.

The information of HCC genes is abundant in the gene expression profile. Over the past two decades, more and more high-throughput techniques have been developed to do the genome-wide analysis of gene expression and their interactions [2]. The gene co-expression network of HCC provides a tool for studying the gene regulation based on gene expression data, which provides favorable conditions for finding more stable biomarkers and gene targets.

Many investigators have made preliminary progress in the search for biomarkers of HCC. Through the study of HCC samples, the researchers have found a number of biomarkers which have significant changes in the expression, such as alpha-fetoprotein (AFP). Tada et al found that the percentage of AFP-13 is strongly related to the staging and size of HCC [3]. These tasks have much significance to clearing pathogenesis of HCC. At present, analysis of gene co-expression network has been widely used in complex disease research. Gene network is a complex dynamic system. In order to achieve genetic relationship mining, we need to select the community structure from the complex network, which is called gene co-expression module. The genes within these modules share similarities in physiology and gene function [4]. Therefore, analyzing the co-expression network of HCC and enriching the gene modules with similar function by clustering can help us to deeply understand the gene interaction in the process of HCC, so as to get the disease-related gene modules and comprehend its pathological mechanism systematically.

Methods

Process of Research

As shown in Figure 1 we first obtained gene expression data from the public database of HCC, which contains 20,673 genes in the sample data. After the data were normalized, 328 genes closely related to HCC were selected on the basis of Student's Test and correlation analysis. Then, the co-expression network of different genes was constructed according to the relativity of these genes, and the network was clustered to divide the network into several functional gene modules. Based on the systematical study of these gene modules, the functional modules of the selected genes were analyzed for enrichment and further exploration of genes related to HCC. Next, two candidate biomarkers were identified. Finally, the support vector machines (SVM) machine learning method was used to evaluate the disease classification of the selected gene modules and candidate biomarkers.

Data Source

HCC gene expression profile data were obtained from Gene Expression Omnibus (GEO) database of the National Center of Biotechnology Information (NCBI). Its number is GSE20948, containing 20,673 genes in 14 experimental samples and 14 control samples. The experimental samples and the control samples correspond to each other, named from GSM523800 to GSM523827. Fourteen experimental samples among them are

infected cells. They were infected with the genotype 2a HCV clone, JFH-1 at a multiplicity of infection (MOI) of 3. At 6, 12

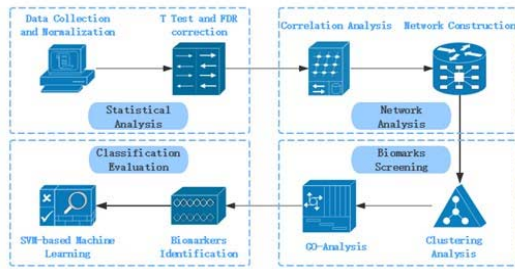


Figure 1- Map of research process

18, 24 and 48 hours' post-infection, cellular RNA was extracted and the gene expression value was measured from the gene expression profile. Another fourteen control samples were normal cells without HCV infection under the same conditions as the experimental group.

Overview of Research Methods

Statistical Analysis

The expression data are preprocessed for normalization with the gene expression console software. We apply Student's Test to achieve the initial screening of differentially expressed genes on a large scale. Using the R language as the data processing tool, we calculate the P value of two vectors of the experimental group and control group. The smaller the P value, the greater difference between the two vectors, and the greater difference in gene expression between the infected samples (disease state) and the non-infected samples (healthy state). Then False Discovery Rate (FDR) correction is performed by using the `p.adjust` function to obtain the FDR corresponding to each gene data. The significance of FDR is the proportion of false conclusions that are claimed to be significant. The goal is to control the false discovery rate below a certain value.

Gene Co-expression Network Analysis and Clustering

We use Pearson Correlation to calculate the Pearson Correlation Coefficient (PCC) of the expression vector between different genes to identify the first 1% of the genes with the greatest difference in disease and health status [5]. We then use Cytoscape to map all co-expressed gene pairs into a gene co-expression network (Figure 1).

In the co-expression network, the nodes represent the genes, and the edges represent the co-expression interactions [6]. The information can be transferred smoothly on these edges. A collection of nodes means that a set of co-expressed genes collectively carries out certain functions. In this paper, Markov clustering algorithm (MCL) is used to cluster [7]. In order to get the appropriate number of community sets, we choose the Granularity Parameter of 1.8 and the other parameters of the default parameters. After the clustering algorithm is implemented, the clustering results are represented by a number of modules, each of which contains a number of genes.

Identifying Candidate Biomarkers by GO-Analysis

For the four gene modules with the highest correlation, the functional analysis of the gene network is carried out through Network Ontology Analysis (NOA) method and Gene Card website [8]. The NOA online website provides functional gene analysis for free which gives static and dynamic analysis for gene networks. Through the functional analysis of genes, we

identify two important genes, which are defined as candidate biomarkers.

Determine the Accuracy of Classification by SVM

Our candidate biomarkers are determined by our functional enrichment analysis of the genes in the gene modules. In order to further confirm the relevance of HK2, KLF4 and HCC and the accuracy of the classification, we use SVM to evaluate the effect of the classification of disease for the four most relevant gene modules. We choose the method of Leave-one-out to classify the existing data and make the ROC curve to get the accuracy index of classification [9].

Results

Screening of Differentially Expressed Gene

The `t.test()` function is used to test the data and get the P value of each gene data. According to the size of the P value in ascending order, we pick out the smallest P value of the first 500 genes as candidate genes. All the 500 genes selected with the smallest P value meet the required FDR range (FDR < 5%). We list the first ten genes and their P value and FDR in Table 1.

Table 1- P value and FDR of the selected genes

Gene Name	p-value	FDR
PDGFRA	6.28E-16	1.30E-11
PRR3	5.65E-13	5.84E-09
NFKBIZ	1.02E-12	7.00E-09
SYNGR2	4.19E-12	2.16E-08
RCAN3	5.84E-12	2.41E-08
TUBB6	7.99E-12	2.43E-08
SLC25A33	8.24E-12	2.43E-08
KLF2	1.43E-11	3.68E-08
CRIP3	1.76E-11	4.03E-08
PHYHIP1L	2.20E-11	4.47E-08

Correlation Analysis

In the R software, we use the `cor()` function to obtain the PCC matrix of two pairs of genes to take out the correlation coefficient values of disease and health state and subtract the difference to get the absolute value of the relative PPC, which ranges between 0 to 2. There are 124750 values and the size of the difference value means the strength of the correlation between the two genes. Additionally, we use `hist()` function to make the frequency distribution and use the `lines()` function to make the fitting curve (Figure 2). The horizontal axis is the absolute value of the relative PPC between 0 and 2, and the vertical axis is the number of occurrences of this value.

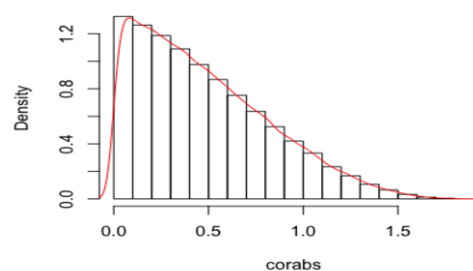


Figure 2- Frequency distribution of the correlation

Figure 2 shows that the greater the absolute value of the relative PPC, the smaller the data, that is, the weaker the correlation between the two genes. Therefore, with the frequency distribution map, the absolute value of the relative PPC is arranged in descending order and the maximum 1% of the data is retained. The first 1200 pair pairs are screened and their absolute values of the relative PPC are obtained. The first twelve gene pairs are chosen to be listed in Table 2.

Table 2- Absolute values of the relative PPC of gene pairs

Gene1 Name	Gene2 Name	Cor (abs)
DTX3L	PLEKHA2	1.841587988
SLC7A2	SMAD6	1.828728355
FGG	IRF9	1.826512894
RPF2	SMAD6	1.814375402
IRF9	RCAN2	1.81315706
RAB27B	MAFF	1.806385076
FGG	SMAD6	1.805861711
IRF9	PLEKHA2	1.791235586
TMEM99	PLEKHA2	1.790806898
RCAN2	DTX3L	1.790245603
VCAN	PSAT1	1.785278759
BICD2	RASSF5	1.783599316

Construction of Gene Coexpression Network and Clustering Analysis

To explore the co-expression changes induced by the highly relevant gene pairs, the 1200 pairs of genes selected above are mapped into the gene co-expression network with the help of the Cytoscape software (Figure 3). Each vertex in the graph represents a gene. When the co-expression of the two genes (ie, the absolute value of the relative PPC) is greater than the selected threshold, an edge is connected to indicate the significant correlation between them.

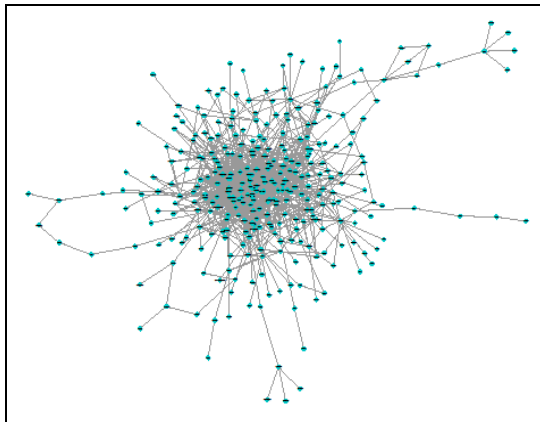


Figure 3 - Gene co-expression network

GO-Analysis

We select the four gene modules with the strongest correlations and analyze them with the NOA online website to get the functional analysis of the genes within the modules (Table 3). In combination with the analysis of Gene Card website, we identify HK2 and KLF4 as potential candidate molecular biomarkers.

Table 3- Result of GO-Analysis

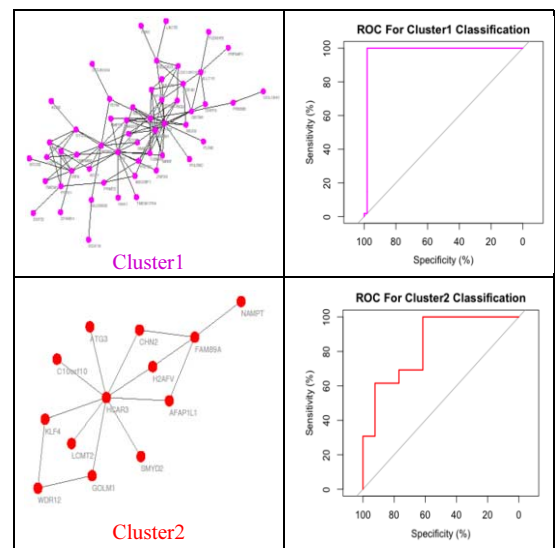
GO: term	p-value	Corrected p-value	Term name
GO:0000460	9.4E-4	0.1551	Maturation of 5.8S rRNA
GO:0000466	9.4E-4	0.1551	Maturation of 5.8S rRNA from tricistronic rRNA transcript
GO:0000676	9.4E-4	0.1551	Nicotinamide metabolic process
GO:0010999	5.4E-4	0.1022	Regulation of eIF2 alpha phosphorylation by heme
GO:0045993	5.4E-4	0.1022	Negative regulation of translational initiation by iron
GO:0046984	5.4E-4	0.1022	Regulation of hemoglobin biosynthetic process
GO:0046986	5.4E-4	0.1022	Negative regulation of hemoglobin biosynthetic process

SVM-based machine learning method

Through the systematic study of these functional modules and candidate biomarkers, we use the method of Leave-one-out to do the machine learning of the existing data in the R software and get the corresponding value of accuracy (AUC) (Table 4) and ROC curves (Figure 4).

Table 4- Number and AUC of selected gene modules

Module	Gene Number	Auc
Cluster1	54	0.98
Cluster2	13	0.84
Cluster3	12	0.97
Cluster4	11	0.98



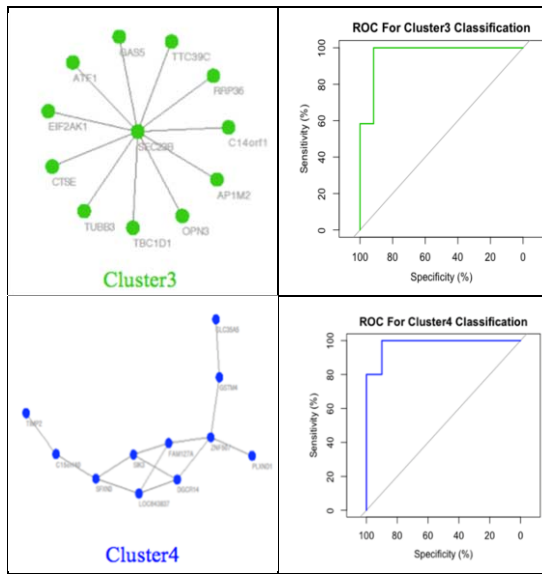


Figure 4- Gene modules and corresponding ROC curves

From Table 4 and Figure 4, we can see that the SVM classifier has a good AUC of 0.98, 0.84, 0.97 and 0.98, which are very close to 1 after training and learning for the four gene modules. And the corresponding ROC curves are very close to the upper left corner of the highest threshold of the specificity and sensitivity. It can be seen that the SVM classifier shows good performance in the classification of the four gene modules, indicating that the four gene modules can distinguish disease state and health state. And the gene HK2 and gene KLF4 contained show significant differences in the two states and most likely can be identified as biomarkers of HCC.

Discussion

Looking for biomarkers of HCC and exploring the molecular mechanism of HCC are of great significance for early prediction and early treatment of HCC. HK2 is a Protein Coding gene which is associated with the occurrence of some tumors. Among its related pathways are Regulation of Glucokinase by Glucokinase Regulatory Protein and Translation Insulin regulation of translation. KLF4 is a tumor suppressor gene which participates in cell differentiation, cell proliferation, necrosis and angiogenesis. The encoded protein is thought to control the G1-to-S transition of the cell cycle following DNA damage by mediating the tumor suppressor gene p53. These two genes are to some degree, related with the occurrence or metastasis and invasion of the tumor and cell differentiation, angiogenesis and so on. This also confirms the reliability and accuracy of our approach from another side.

Compared with the previous studies, the molecular mechanism based on HCC in this paper should be a hypothesis of a network of gene dynamic regulation, and more consideration is given to the potential interconnections between genes. Through the construction of co-expression network to further explore the molecular mechanism of liver cancer, the method of mining effect should be more reasonable. Although the correlation between the two genes and the occurrence of HCC needs to be confirmed by further experiments, it is believed that the applicability of this method will be further demonstrated with

the further accumulation and perfection of high-throughput experimental data.

The occurrence and development of cancer are caused by the mutual disturbance of multiple functional pathways. Therefore, during the occurrence and development of HCC, tumor-associated genes should be a dynamic process of mutual regulation of the network. Consequently, we analyze the microarray data of HCC with the combination of traditional bioinformatics analysis method and the increasingly active system biology analysis and use bioinformatics methods such as normalization, statistical analysis, clustering analysis, GO-analysis, SVM and so on. The construction of gene co-expression network is completed and the network is searched for genes related to HCC. As a result, the gene HK2 and gene KLF4 are obtained and the functional modules of these two genes are proved to possess a great ability of classification by using SVM, which will contribute to the future detection and personalized medicine of HCC. This will be of great help and significance for the development of precision medicine to improve the lives and health of patients.

Acknowledgements

This work was supported by Chinese National High-tech R&D Program (2015AA020109), National Key Scientific Instrument and Equipment Development Project (2016YFF0103200), the Fundamental Research Funds for the Central Universities of China, National Natural Science Foundation of China (61572287) and Shandong Provincial Natural Science Foundation, China (ZR2015FQ001).

References

- [1] D. Joshi, C. Taylor, J. Maggs, M. Heneghan, K. Childs, I. Carey, N. Heaton, J. O'Grady, K. Agarwal, and A. Suddle, Hepatocellular carcinoma (HCC) in HIV-positive patients: a more aggressive disease course?, *HIV MEDICINE* **12** (2011), 36-36.
- [2] W. Zhao, P. Langfelder, T. Fuller, J. Dong, A. Li, and S. Hovarth, Weighted Gene Coexpression Network Analysis: State of the Art, *Journal of Biopharmaceutical Statistics* **20** (2010), 281-300.
- [3] T. Tada, T. Kumada, H. Toyoda, S. Kiriya, Y. Sone, M. Tanikawa, Y. Hisanaga, S. Kitabatake, T. Kuzuya, K. Nonogaki, J. Shimizu, A. Yamaguchi, M. Isogai, Y. Kaneoka, J. Washizu, and S. Satomura, Relationship between Lens culinaris agglutinin - reactive α - fetoprotein and pathologic features of hepatocellular carcinoma, *Liver International* **25** (2005), 848-853.
- [4] S. Zhang, H. Zhao, and M. Ng, Functional Module Analysis for Gene Coexpression Networks with Network Integration, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* **12** (2015), 1146-1160.
- [5] M. Rahman and Q. Zhang, Comparison among pearson correlation coefficient tests, *Far East Journal of Mathematical Sciences* **99** (2016), 237-255.
- [6] Z.P. Liu, Identifying network-based biomarkers of complex diseases from high-throughput data, *BIOMARKERS IN MEDICINE* **10** (2016), 633-650.
- [7] S. Van Dongen and C. Abreu-Goodger, Using MCL to extract clusters from networks, in, United States, pp. 281-295.
- [8] S. Zhang, J. Cao, Y.M. Kong, and R.H. Scheuermann, GO-Bayes: Gene ontology-based overrepresentation analysis using a Bayesian approach, *Bioinformatics* **26** (2010), 905-911.
- [9] W. Mao, X. Mu, Y. Zheng, and G. Yan, Leave-one-out cross-validation-based model selection for multi-input multi-output support vector machine, *Neural Computing and Applications* **24** (2014), 441-451.

Address for correspondence

Corresponding author: Jing-song Li.

Address: College of Biomedical Engineering and Instrument Science,
Zhejiang University, 38 Zheda road, Hangzhou 310027, China.
Email: ljs@zju.edu.cn. Tel.: +86-571-87951564

Corresponding author: Zhiping Liu.

Address: College of Control Science and Engineering, Shandong
University, 17923 Jingshi road, Jinan 250061, China.
Email: zpliu@sdu.edu.cn. Tel.: +86-137-9113-2032

An Interactive Platform to Visualize Data-Driven Clinical Pathways for the Management of Multiple Chronic Conditions

Yiye Zhang^a, Rema Padman^b

^a Division of Health Informatics, Department of Health Policy and Research, Weill Cornell Medical College, New York, NY, USA,

^b The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

Patients with multiple chronic conditions (MCC) pose an increasingly complex health management challenge worldwide, particularly due to the significant gap in our understanding of how to provide coordinated care. Drawing on our prior research on learning data-driven clinical pathways from actual practice data, this paper describes a prototype, interactive platform for visualizing the pathways of MCC to support shared decision making. Created using Python web framework, JavaScript library and our clinical pathway learning algorithm, the visualization platform allows clinicians and patients to learn the dominant patterns of co-progression of multiple clinical events from their own data, and interactively explore and interpret the pathways. We demonstrate functionalities of the platform using a cluster of 36 patients, identified from a dataset of 1,084 patients, who are diagnosed with at least chronic kidney disease, hypertension, and diabetes. Future evaluation studies will explore the use of this platform to better understand and manage MCC.

Keywords:

Critical Pathways; Computer Graphics; Multiple Chronic Conditions

Introduction

Multiple chronic conditions (MCC) have significant health management and cost implications worldwide. For example, in developed nations, one in four people have multiple chronic conditions[1; 2]. In the US, over one-third of beneficiaries of Medicare, the US single-payer, social insurance program, have at least 4 concurrent chronic conditions, and they alone are responsible for 74% of total Medicare spending[3]. Yet, previous literature has suggested that there is still a gap in efficient care coordination of these patients, as commonly used clinical practice guidelines (CPGs) are focused on the optimal care of each condition and not on the composite of many conditions [4]. In this study, we report on the design and development of a prototype of an interactive visualization platform, based on information technology (IT) and advanced data-analytics, which captures and displays how MCC patients are treated in real settings, given their varying demographics and concurrent conditions. This approach has the potential to facilitate shared understanding and exploration of possible disease and treatment pathways and support shared decision making between clinicians and patients.

Background

Health IT has gained widespread recognition in facilitating effective management of patients through electronic health records (EHR) systems and their various functions in supporting clinical decision support (CDS) [6]. Clinical pathways are multidisciplinary care plans to detail essential care delivery steps that can be embedded in the EHR, and are expected to link evidence to practice and optimize clinical outcomes while maximizing clinical efficiency [5]. In this study, we place an emphasis on the fact that our clinical pathways are learned from EHR data and display data-driven clinical pathways of MCC. The learning algorithm draws on techniques from machine learning, statistics, and operations research, and combines it with innovative data structuring and modeling to handle the multiple dimensions of patient care data.

In order for these data-driven clinical pathways to truly contribute to learning health systems, we must address another significant challenge, which is the usability of the clinical pathway learning algorithm and its outputs for multiple stakeholders in healthcare. Results from previous work [6; 7] are presented as pathways of complex data representations, which require a sophisticated level of understanding of informatics and analytics to interpret and use. Therefore, this study aims to design and implement a prototype of an interactive, visual analytics clinical pathway platform that embeds the current methods and meets the information needs of diverse users, including healthcare professionals, researchers, and patients, without requiring such specialized knowledge. Visual analytics utilizes the high bandwidth processing capabilities of the human visual system to more efficiently support interactive data exploration, analysis, and absorption of complex information [8; 9]. Through this research, we expect that (a) managers of healthcare practices, such as an accountable care organization (ACO), can easily review clinical pathways of local practices to determine variations, exceptions, and optimal care patterns; (b) clinicians and patients are provided with resources to encourage shared-decision making at the point of care; and (c) researchers can generate new hypotheses for targeted randomized clinical trials (RCT) through identification of promising pathway patterns for which strong, scientifically-tested evidence does not currently exist.

Methods

Technology Components of the Visualization Platform

The backend system of the visualization platform is based on a clinical pathway learning algorithm described in Zhang et al [6]. The algorithm particularly focuses on the co-progression of multiple concurrent conditions and associated interventions, and thus is a natural choice for the platform in modeling the clinical pathways of MCC management. Given as input the EHR data of a patient population that includes chronological information on encounter types, diagnoses, and interventions, all recorded at different points in time, the algorithm performs patient clustering, followed by Markov chain modeling, in learning the dominant patterns of clinical pathways within each patient cluster. Figure 1 displays a schematic representation of the algorithm.

As shown in Figure 1, Zhang et al used a dimension reduction approach (Step (a) in Figure 1) to summarize multiple aspects of information at each point in time. In other words, each node in the pathway, once visualized as in the output in Figure 1, is in a complex, coded-form that may be quite incomprehensible to readers of the clinical pathways. Therefore, creating an interactive visualization platform to decode this complexity is a natural extension of the study, such that users without background in clinical pathway learning can easily use and understand the outputs for management of patient populations as well as individual patients. While previous work has developed visualization capabilities for clinical pathways [8; 10; 11], our pathway learning algorithm and visualization platform distinguishes itself by particularly focusing on the need to track MCC and associated interventions over time for targeted patient subgroups and individuals.

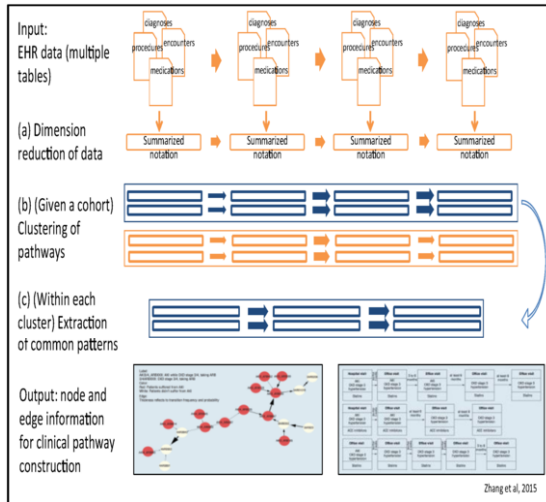


Figure 1 – Outline of Pathway Learning Algorithm

The frontend system of the platform is created using Django (Django Software Foundation), a Python web framework. As demonstrated in later sections, users can input their EHR data and user-defined filtering information for Step (c) in Figure 1, and obtain summary statistics and visualization of the dominant clinical pathways in the identified patient population. Filtering can be done by the transition probability and frequency of pathway elements, as described in Zhang et

al [6]. Visualization is created using force-layout form of D3 [12], a JavaScript library. Compared to other libraries such as Gephi [13] and Plotly [14], D3 library contains more interactive, flexible and customizable features such as allowing users to place mouse over a component in the visualizations to facilitate deeper exploration. As an extension to the methods in Zhang et al [6], we also incorporate temporal information in the clinical pathways, so that we can track the duration between clinical visits, and differentiate a pathway showing the progression of CKD stage 3 to stage 4 that took place *within* a month, from another pathway showing progression in *more than* a month, for instance. Current implementation categorizes duration into 2 categories, within and beyond 4 months, and color-coded to mark the difference. However, this categorization can be readily modified depending on the exploration of interest with distinct pathways.

Data used to Demonstrate the Platform

For demonstration, we used a dataset from the EHR of a community Nephrology practice in southwestern Pennsylvania, which contains data on 1,084 patients whose initial diagnoses in 2009 included chronic kidney disease (CKD), hypertension, and diabetes, with additional chronic conditions developing with the progression of the disease over time. The algorithm learned the pathways of these patients, consisting of the first 5 encounters from 2009 to 2016. As variable of interest in the analysis, we particularly track progression of the following diagnoses: CKD stages 1 to 5, end state renal disease (ESRD), acute kidney injury (AKI), hypertension, and diabetes; and medications summarized by their drug classes including Angiotensin-converting-enzyme (ACE) inhibitors, angiotensin receptor blockers (ARB), diuretics, statins, among others. In this data, gender ratio of female to male is 0.42 and 0.58. Over 90 percent of the patients are Caucasian. Table 1 summarizes some descriptive statistics associated with this group of 1,084 patients, including number of diagnoses, medication prescriptions, and hospitalizations.

Table 1 – Descriptive Summary of Data

Number	Average	Max	Min
Diagnosis	3.4	6	3
Medication	2.3	20	0
Hospitalization	0.2	4	0
Months in treatment	14.1	76	0.2

Results

Clustering of patients

Figure 2 shows the clustering results from the above dataset. The clusters were found using hierarchical clustering [6]. As the figure illustrates, the algorithm finds clusters of patients who have distinct characteristics, such as those with no changes in their disease over time, or those who develop specific types of complications. The clusters are color-coded in Figure 2. In the demonstration study, a cluster of 36 patients who developed acute kidney injury (AKI) is extracted from the full dataset for further analysis and visualization. AKI is a serious complication of CKD that can result in rapid deterioration of the kidney functions and even mortality. Table 2 summarizes some descriptive statistics associated with this group of 36 patients for comparison with the whole group.

Visualization of this cluster was created under a filtering of at least 3 visits for pathway length, at least 0.3 for transition probability, and at least 1 for transition frequency. These filters are described in [6].

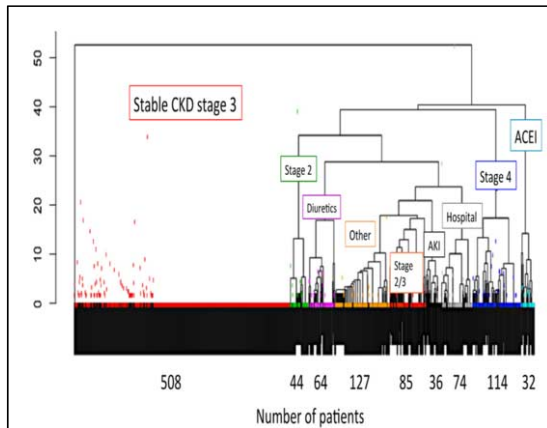


Figure 2 - Clustering of patient groups. Color indicates clusters

Table 2- Descriptive Summary of Data within AKI Cohort

Number	Average	Max	Min
Diagnosis	4.3	6	4
Medication	2.6	8	0
Hospitalization	0.2	2	0
Months in treatment	7.1	18.9	0.6

Visualization of the AKI cluster

Figure 3.1 shows a screenshot of visualized clinical pathways from one of the clusters identified above. Pathways originate from the blue node on the right hand side, and diverge into different directions depending on patient and treatment characteristics. Each node in the visualization represents a clinical visit, capturing information on encounter types, diagnoses, and medications prescribed from each visit. Users can hover over the node and see the exact information encoded. For example, the node highlighted in Figure 3.1 is a visit of Office (O) type, where diagnoses of AKI, CKD stage 3 (CKD3), diabetes (DM), and hypertension (HP) were noted. Vitamins were prescribed in the visit. Sizes of the nodes are relative to the number of edges it is connected to. In addition, users can drag nodes to desired locations for enhanced interpretability.

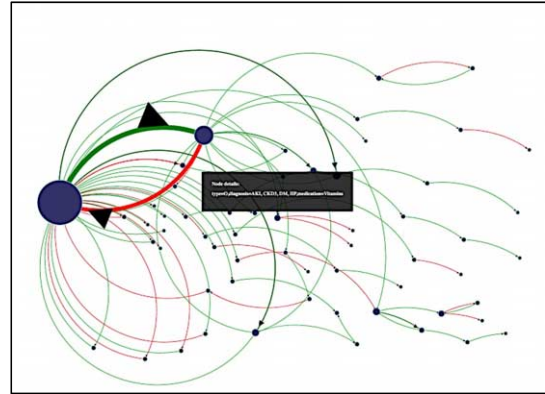


Figure 3.1 - Visualized clinical pathways: node highlighted is a visit of Office (O) type, where diagnoses of AKI, CKD stage 3 (CKD3), diabetes (DM), and hypertension (HP) was noted. Vitamins were prescribed.

Further, the edges and arrow sizes between each node capture multiple categories of information about the transitions between clinical visits, particularly through color-coding and thickness. Green and red links indicate less than 4 months, and at least 4 months, respectively, of gap between visits, as explained in the legend on the right-hand side (Figure 3.2). Thickness of the edges reflects the probability of transitions; the thicker the edges, the more likely that the transitions are observed in the data, such as displayed by the thick, red edge in Figure 3.2.

A search function is built into the platform that allows users to identify visits with certain criteria. For example, Figure 3.2 shows a search for visits where patients' diagnoses include CKD stage 3, and medication prescriptions include vitamins. The search bar on the top left indicates that there are 6 clinical visits that match with the specified criteria, and information of one is highlighted. The search feature can be expanded by adding more functionalities or categories as needed.



Figure 3.2 - Search function highlighting interesting visits

Using the Visualization Platform

Figure 4.1 - Input form for specifying critical filters

To use the platform, users need to enter the path to their data in JSON (JavaScript Object Notation) format, which contains each patients' clinical events and their timestamps over a period of interest. If necessary, users can filter clinical pathways by their commonality. Currently implemented thresholds are the number of time points in the pathways, transition probability between visits, and number of transitions. For example, Figure 4.1 displays the input form with a filter, where only pathways with at least 3 time points, transition probability of at least 0.1 and transition frequency of at least 2 are included in the analysis. Figure 4.2 shows the updated pathway visualization, now much simpler compared to the initial visualizations in Figures 3.1 and 3.2 with the original filter. The resulting pathways that are visualized can be shorter than 3 time points due to the requirement on transition probability and frequency.

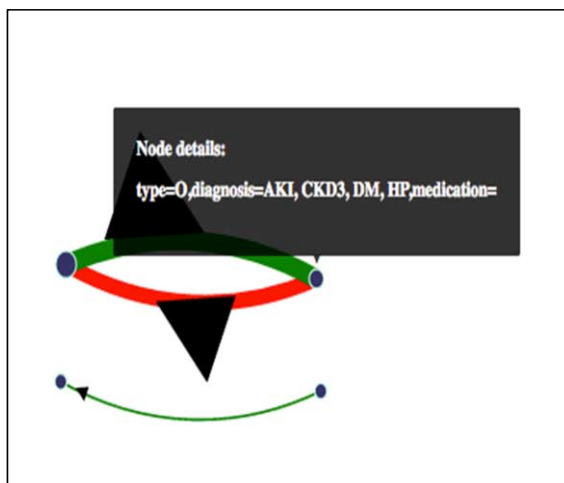


Figure 4.2 - Visualized clinical pathway after applying filters

Displaying Clinical Pathway Statistics

source	target	weight	count	count_source	count_target
V13	V29	0.571428571429	12	21	0
V29	V29	0.394366197183	28	71	0
V29	V13	0.211267605634	15	71	0
V47	V30	0.666666666667	2	3	0

Figure 5.1 - Relevant statistics of the visualized pathways

Number of occurrences of each node, transition probability and frequency of each edge, and other statistics are also available for display in the platform as shown in Figure 5.1. “Source” indicates the originating visit and “target” is the transitioned visit. For example, in the pathway visualization in Figure 3, we see from Figure 5.1 that there are 15 instances among the 36 patients' pathways where patients develop an episode of AKI (row 3). Figure 5.1 also shows that this transition, an office encounter with a hypertensive, diabetic patient in CKD stage 3 (source: V29), progressing to develop an episode of AKI (target: V13), has a transition probability of 0.21. The corresponding code information for the visit IDs can be found by clicking the “See Node/VVNode Details” button in the platform. Figure 5.2 shows the information of 7 nodes including V13. While not shown in the paper due to space limitations, visit IDs can be displayed as part of the mouse-over function to show node details, to allow easy mapping of the pathway elements with corresponding statistics.

visit ID	visit type	diagnosis	procedure	medication
V0	Otype	AKI CKD3 DM HP	-	Benzodiazepine anticonvulsants . Miscellaneous anemetics . Statins . Sulfonylureas .
V1	Otype	CKD3 DM HP	-	Adrenergic bronchodilators . Benzodiazepine anticonvulsants . Benzodiazepines . Glucocorticoids . Miscellaneous anemetics .
V10	Otype	AKI CKD3 DM HP	-	Angiotensin II inhibitors with thiazides . Gamma-aminobutyric acid analogs .
V11	Otype	CKD3 DM HP	-	Antacids . Antipruritic agents . Antihypertensive agents . Minerals and electrolytes . Purine nucleosides . Urinary pH modifiers .
V12	Otype	AKI CKD3 DM HP	-	Angiotensin II inhibitors with thiazides . Vasodilators . Vitamins .
V13	Otype	AKI CKD3 DM HP	-	-
V14	Otype	CKD3 DM HP	-	-

Figure 5.2 - Look up table of the visit IDs shown in Figure 5.1

Discussion

This study presents a prototype visualization platform that can potentially be used for other types of data with simple and easy modifications, provided the format of the input data remains the same and there are a comparable number of nodes. One challenge that we recognize is that the structure may become unstable when there are many nodes. In this case, force-layout graphs cannot be clearly displayed. This problem likely needs to be solved via the backend algorithm, which must explore options to summarize clinical pathways in a scalable and clinically valid fashion.

Evaluation of this platform with actual users in a formal experimental setting, such as through think-aloud protocol and focus groups, is an important future work. To accommodate this task, we are working on adding human computer interaction (HCI) capabilities, such as recording the number of times a node is clicked by a user in order to understand what features users are most interested in. In addition, prediction of future states in the clinical pathways is also a problem of interest, as well as managing not only the clinical aspects but also the costs of interventions. Based on our follow-up studies on prediction of outcomes, and overlaying the medical costs, following a particular pathway [15; 16], we plan to add functions for the predictive modeling and cost pathway mapping to the platform as well.

Conclusion

This paper describes an extension of our existing clinical pathway learning algorithm by developing an interactive visualization platform, which displays the co-progression of multiple concurrent disease conditions and their associated interventions such as medications. Particularly focused on use cases in a specialty care setting for patients with MCC, but equally applicable to primary care, we aim to show that the visualization platform, using information technology and data analytics, has the potential to be easily used and interpreted by diverse end users, such as clinicians, patients and researchers, for patient education and empowerment as well as for shared decision making and developing generalizable research and insights. The platform has functions for personalized filtering and searches, allowing users to explore clinical pathways of specific interest to their condition. Future work will include formal experiments as evaluation studies, and enhancement of the platform usability in terms of backend learning algorithm and added HCI capabilities.

Acknowledgements

We are grateful to the physicians and staff of the community nephrology practice who generously provided data from their Electronic Health Record for this study. We particularly thank Dr. Teredesai, MD, Dr. Xie, MD, PhD, and staff, L. Smith and A. Barletta, who gave us important clinical and technical information about the data and the key characteristics of CKD, AKI and the treatments. We also thank the two graduate students, Rohit Bandi and Ahana Sheshadri, who helped us with the implementation of the visualization platform.

References

- [1] K. Barnett, S.W. Mercer, M. Norbury, G. Watt, S. Wyke, and B. Guthrie, Epidemiology of multimorbidity and implications for health

- care, research, and medical education: a cross-sectional study, *Lancet* 380 (2012), 37–43.
- [2] U.D.o. Health and H. Services, Multiple chronic conditions—a strategic framework: optimum health and quality of life for individuals with multiple chronic conditions, Washington, DC: US Department of Health and Human Services 2 (2010).
- [3] C.f. Medicare and M. Services, Chronic conditions among Medicare beneficiaries, chartbook, 2012 edition, Baltimore, MD (2012).
- [4] A.L. Leppin, V.M. Montori, and M.R. Gionfriddo, Minimally Disruptive Medicine: A Pragmatically Comprehensive Model for Delivering Care to Patients with Multiple Chronic Conditions, *Healthcare (Basel)* 3 (2015), 50–63.
- [5] T. Rotter, L. Kinsman, E. James, A. Machotta, H. Gothe, J. Willis, P. Snow, and J. Kugler, Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs, *Cochrane Database Syst Rev* (2010), CD006632.
- [6] Y. Zhang, R. Padman, and N. Patel, Paving the COWpath: Learning and visualizing clinical pathways from electronic health record data, *J Biomed Inform* (2015).
- [7] Y. Zhang, R. Padman, and L. Wasserman, On Learning and Visualizing Practice-based Clinical Pathways for Chronic Kidney Disease, in: *American Medical Informatics Association 2014 Annual Symposium*, Washington, DC, 2014.
- [8] J.S. Hirsch, J.S. Tanenbaum, S. Lipsky Gorman, C. Liu, E. Schmitz, D. Hashorva, A. Ervits, D. Vawdrey, M. Sturm, and N. Elhadad, HARVEST, a longitudinal patient record summarizer, *J Am Med Inform Assoc* 22 (2015), 263–274.
- [9] J.J. Thomas, Illuminating the path:[the research and development agenda for visual analytics], *IEEE Computer Society*, 2005.
- [10] D. Klimov, A. Shknevsky, and Y. Shahar, Exploration of patterns predicting renal damage in patients with diabetes type II using a visual temporal analysis laboratory, *J Am Med Inform Assoc* 22 (2015), 275–289.
- [11] A. Perer, F. Wang, and J. Hu, Mining and exploring care pathways from electronic medical records with visual analytics, *J Biomed Inform* 56 (2015), 369–378.
- [12] M. Bostock, V. Ogievetsky, and J. Heer, D³ data-driven documents, *IEEE transactions on visualization and computer graphics* 17 (2011), 2301–2309.
- [13] M. Bastian, S. Heymann, and M. Jacomy, Gephi: an open source software for exploring and manipulating networks, *ICWSM* 8 (2009), 361–362.
- [14] A. Rohatgi, WebPlotDigitizer, URL <http://arohatgi.info/WebPlotDigitizer/app> (2011).
- [15] Y. Zhang and R. Padman, Innovations in chronic care delivery using data-driven clinical pathways, *Am J Manag Care* 21 (2015), e661–668.
- [16] Y. Zhang and R. Padman, Data-Driven Clinical and Cost Pathways for Chronic Care Delivery, *American Journal of Managed Care* (2016).

Address for correspondence

Yiye Zhang, 425 E. 61st Suite 301, New York, NY 10065.
yiz2014@med.cornell.edu

III. Human, Organizational, and Social Aspects

This page intentionally left blank

Developing Digital Health and Welfare Services in an International Multidisciplinary Student Team

Outi Ahonen^{a,c}, Elina Rajalahti^a, Jonas Tana^b,
Gun-Britt Lejonqvist^b, Ulla-Mari Kinnunen^c, Kaija Saranto^c

^a Research, development and innovation unit, Laurea University of Applied Sciences, Espoo, Finland

^b Department of Health and Welfare, Arcada University of Applied Sciences, Helsinki, Finland

^c Department of Health and Social Management, University of Eastern-Finland, Kuopio, Finland

Abstract

The rapid growth of digital health and welfare services demands new competences for health and social care, information technology, and business professionals. This study aims to describe the competences that students have before their studies and those they expect to gain from the study module "Developing Digital Health and Welfare Services" in multiprofessional groups during their bachelor studies. This study reports open-ended questions about students' knowledge concerning digital health prior to the study units. The results, analyzed by QSR NVivo 10 for Windows, show that students are keen to learn about developing digital health and welfare services, and they see that multiprofessional work requires a communicative environment and respect for every profession. Students also believe that they have competences to bring to the multiprofessional group. A successful multidisciplinary development of digital health and welfare services requires changes and cooperation in education between various professions.

Keywords:

Informatics; Education; Professional Competence

Introduction

Wellbeing, health promotion, disease prevention, and citizen's empowerment to his/her own life are the core areas of social and health politics. Health informatics plays a crucial role in accomplishing these elements [1]. New digital health services are essential in order to improve care all over the world, to increase the level of patients' engagement in their own care, to arrange quality health services with universal access, and to create a sustainable financial ground for health care systems [2]. To achieve effective implementation, the customer-centric service culture in health care requires a human-centered design approach to co-creation innovation and also skills, such as sensitivity and attitudes [3]. Well-functioning eHealth services need new competences from professionals and citizens [4]. Nearly half of the European Union (EU) population (47%) does not have proper digital skills, yet in the near future, 90% of jobs will require it. The EU Commission supports efforts to enhance the digital skills and qualifications of the population and to increase the level of ICT professionalism [5]. An effort to achieve this has been made through the European Computer Driving License (EDCL), which since 1995, has given a worldwide format to provide general knowledge and skills about ICT to all professionals on different educational levels [6]. The biomedical and health informatics standardized curriculum developed by the International Medical Informatics Association (IMIA) is widely

known [7]. Also, nursing informatics has been part of nursing education for many years [8–11]. Moreover, social workers are required to be competent in the use of informatics [12]. However, the research still shows that students need to increase their knowledge, skills, and competences in informatics [10]. The European Qualifications Framework (EQF) defined by the EU is the general framework of vocational qualifications for competence in eight different levels. A bachelor's level is defined as college completion level 5, and it is the first level of higher professional education. The University of Applied Sciences (UAS) is level 6. The EQF defines competences, skills and knowledge related to all degrees [13], and the directive describes professional minimum competences [14].

This study aims to describe the competences that students have before studies and those they expect to gain from the study module "Developing Digital Health and Welfare Services" in multiprofessional groups. This study is part of an EU-funded project. The final aim of the whole project, called the "Developer of Digital Health and Welfare Service (DeDiWe)", is to create a 30-credit module curriculum based on IMIA's recommendations about Developing eHealth and Welfare Services, and to define and describe the eHealth developers' previous knowledge, skills, and competences in bachelor's degree programs for EQF levels five and six.

The research questions are as follows: 1) What kind of competences do bachelor's students expect to get from participating in the study module for developing digital health and welfare services? 2) What do the bachelor's students experience as most important in multiprofessional work in developing digital health and welfare services? 3) What kind of competences do the students from the different professions bring with them to the multiprofessional group?

Methods

The questionnaire

Data for this study were collected with the help of an e-questionnaire, which was based on the IMIA's curriculum content [7] for EQF levels 5 and 6 [13] and described the user's information technology (IT) levels in relation to the IMIA's curriculum content. The questionnaire is cross mapped with the EDCL's [6] and IMIA's [7] contents. The service design, which is part of the DeDiWe module's curriculum, is included as an addition. Competences for design thinking have been used as a framework [15] to describe the part of the questionnaire relating to the service-design process. The DeDiWe curriculum also is cross mapped with the IMIA's content-based competence evaluation measurement tool. This study reports on three open-ended questions about the students' knowledge of digital health

prior to the study units. This study is part of the students' competence evaluation before the DeDiWe studies.

Data collection and analysis

Data were collected using an e-questionnaire answered by students who enrolled in the course in Fall 2016. The course outline, aims, learning outcomes, and the timetable were introduced to students before the data collection. Answering the questionnaire was voluntary-based. Students were informed to answer based on their own preconceptions and expectations. The e-questionnaire was distributed to all participating students through the eLearning platform used for the study unit. Research data was anonymized.

The questionnaire was sent to all students ($n = 82$) from the European partner schools in Finland, Latvia, and Estonia that participate in the study units and thesis work. Students' answers from the e-questionnaire came in an Excel spreadsheet, which was transformed into a Word format, and read several times by OA and G-BL, before being loaded into NVivo. QSR International's NVivo 10 for Windows software for qualitative data analysis was used for arranging and analysis of the open-ended questions, finding themes, coding data, and calculating word frequency. All the obtained data was screened and structured to seek the most relevant information. NVivo 10 created a data file for each question, resulting in three data files. The data was checked by seeking synonymies to the words through a text search query in every questionnaire's data separately, which showed the trends for the inductive classification. It also used theoretical knowledge for the classification. After that, the project researchers chose the items for every node by reading the text and designating the node to the item. Some of the items could be in two nodes. NVivo10 helped by coding stripes to see how the items fitted into the nodes and data. After the nodes' references were chosen, researchers read the text again and designated categories under the nodes by qualitative content analyses [16]. Regarding any overlaps, the researchers OA and G-BL performed the final classification.

Results

The results are preliminary and part of the bachelor's students' competence evaluation results. There were students ($n = 64$) from three countries: Finland ($n = 39$), Latvia ($n = 13$), and Estonia ($n = 12$). The number of students in the different study programs were nursing ($n = 32$), physiotherapy ($n = 3$), biomedical laboratory science ($n = 3$), midwifery ($n = 3$), business administration (BBA) ($n = 4$), BBA-IT ($n = 3$), engineering-IT ($n = 0$), doctoral assistant ($n = 7$), social and welfare ($n = 8$), environmental health ($n = 0$), radiography ($n = 1$), and other ($n = 0$). Of all the students, 76.6% were women and 23.4% were men, 46.9% were ages 19–29, 21.9% were 30–39, 23.4% were 40–49, and 7.8% were over 50. Most of the students (85.9%) were full-time students at the bachelor's level, but some of the students (14.6%) were in the open university. The results are presented in the form of word frequency, text search, and metrics coding queries by using NVivo 10.

There were 376 answers (Table 1). The results are presented by each research question, with the whole data and numerical knowledge about categories and subcategories. Every node has been described with categories, subcategories, and original sentences. In each question, the students were asked to name the three most important aspects from every content category: student expectations, multiprofessional work, and student competences. Not all students gave three alternatives, but only one or two.

Table 1 – Frequencies from all answers in data

Content	First answer	Second answer	Third answer	Together
Students' expectations	49	47	39	135
Multi-professional	46	42	38	126
Students' competences	43	39	33	115
All answers	138	128	110	376

Knowledge and skills students expect to gain

The data showed that “Health” ($n = 32$, 5.4%) and “services” ($n = 23$, 3.9%) were the two most common words in the data. There were many synonyms, such as digital and eHealth. The data were divided into two nodes: Biomedical and health informatics (96 references coded with 69.9% coverage) and Developing services (73 references coded with 54.7% coverage). Figure 1 describes the relationships of the nodes and categories, considering the content that students see as important. All nodes and categories are explained in the following sections.



Figure 1– Knowledge and skills students expect to gain from studies

Developing services

The first node in which students expect to gain more information is “Developing services”, and it is classified into four categories. 1) “Develop knowledge, skills, and competence” is the only category that has additional subcategories. The “Develop” category is spread into three subcategories: a) “Current situation” has 25 items expressing needs to understand the current situation in eHealth and eSocial services, including “advanced technologies in social services”, “artificial intelligence machine learning”, and “a greater understanding of the current innovation projects in e-Health in Europe/world”. Students also expressed needs to understand the multidisciplinary perspective: “understanding of business, technology and healthcare together”; b) The “Develop” subcategory has 25 items about the need to understand the developing perspective, such as “I would understand the concept ‘citizen as customer’ in digital health and welfare services”; and c) “Implementing” has one item: “Implementing”. 2) The “Communication” skills category has 12 items, including “improving communications methods to remove cultural barriers”. 3) The “Gain knowledge for future work” category has eight items, such as “be ahead of current trends to be able to predict coming changes in my profession”. And 4) The “Research knowledge, skills, and competence” category has five items, including “analytical skills”.

Biomedical and health informatics

The second node in which students want to gain knowledge and skills is “Biomedical and health informatics”, which is classified into seven categories. There are variations between contents in these categories. 1) The “Health and social care” category has nine items, and mostly students want to know about health care and human functioning in general, but there are also some specific comments, such as “*specifically learning about the brain*”. 2) The “eHealth and welfare” category has 30 items and is the largest category in this node. Students mostly hope to understand the content, for example, “*understanding of business, technology, and health care together*”, but they also have advanced plans, such as “*I wish to learn more about how to use data analysis to diagnose diseases*”. 3) The “Customer’s point of view” category has 11 items explaining the desire to understand the customer’s orientation, such as “*understanding customer behavior in digital the health and welfare environment*” and “*to know more about patient requirements in the e-platform*”. 4) The “Regulations and quality” category has eight items explaining students’ hopes to gain more knowledge about a safe environment, including “*patient data safety and confidentiality in eHealth*”, but also trust, such as “*gaining customers’/patients’ interest and trust regarding using e-health services*”. 5) The “IT skills” category has nine items, including “*Improve my IT-skills*” and “*to have skills to make tools for digital websites*”. 6) The “Informatics category” has 13 items describing students’ wishes to understand how to work with data, such as “*I know where to find the evidence*” and “*health information management skills*”. 7) The final category, “Future possibilities”, has 15 items, which vary from the personal level to international development. Those include “*ability to use digital appliances and skills in the future working as a nurse*” and “*developing health programs in Finland and abroad*”.

Contents students see as important in multiprofessional work in developing digital services

When the words were counted, “different” ($n = 16$, 3%) and “work” ($n = 13$, 3.5%) were the most common. There were also many synonyms like “patient” and “patients”. The data was analyzed and divided into three categories: Biomedical and health informatics (29 references coded with 25.6% coverage), Communicative culture (78 references coded with 55.86% coverage), and High-quality services (35 references coded with 35.8% coverage). Figure 2 describes the relationships of nodes and categories, considering the content that students see as important in multiprofessional work for developing digital services. All nodes and categories are explained in the following sections.

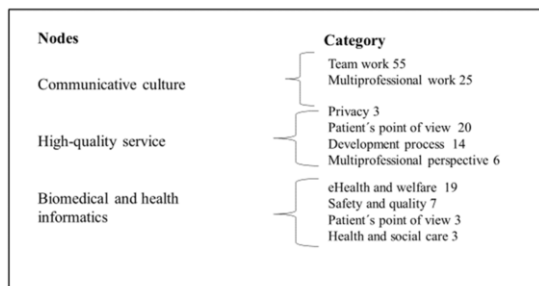


Figure 2 – Content that students see as important in multiprofessional work in developing digital services

Communicative culture

The first node is the “Communicative culture” category, and it was classified into two categories. 1) The “Team work” category has 55 items. Examples are as follows: “*willing to work in a team*”, “*compliance*”, “*dedication*”, and “*coverage*”. 2) The “Multiprofessional work” category has 25 items, such as “*teamwork and collaborating with relevant stakeholder*” and “*understanding other professionals’ roles and competencies*”.

High-quality services

The second node, “High-quality services”, was classified into four categories: 1) “Patient’s point of view” has 20 items, such as “*improve my understanding about the customer in the health and welfare system*”; 2) The “Development process” category has 14 items, including “*development of digital healthcare and welfare services*” and “*leadership and management skills*”; 3) The “Multiprofessional perspective” category has six items, for example, “*every profession has different ideas from different perspectives*”; and 4) The “Privacy” category has three items, including “*privacy issues in health services*”.

Biomedical and health informatics

The third node about the content that students see as important in multiprofessional work in developing digital services is “Biomedical and health informatics”. The node was classified into four categories, and there was a variation between content in the categories: 1) The “eHealth and welfare” category has 19 items, for example, “*Innovation: What works well and what needs to be changed*” and “*research and surveys, monitoring and evaluating digital health and welfare services*”; 2) The “Safety and quality” category has seven items, such as “*data collection while at the same time maintaining confidentiality and privacy*”; 3) The “Health and social care” category has three items, including “*control of people health*”; and 4) The “Patient’s point of view” category has three items, such as “*respecting the right of the patient*”.

Professional competences that students from different branches bring to the multiprofessional group

The most common words in the data were “skills” ($n = 23$, 4.7%) and three words with the same value “understanding”, “work”, and “working” ($n = 9$, 1.8%). There were also many synonyms. The data was analyzed and divided into four nodes. Biomedical and health informatics (17 references coded with 13.1% coverage), Communication (49 references coded with 24% coverage), Health and welfare (41 references coded with 47.5% coverage), and Management and developing (37 references coded with 24% covered). Figure 3 describes the nodes’ and categories’ relationships regarding what professional competences students from different branches bring to the multiprofessional group. All nodes and categories are explained in the following sections.

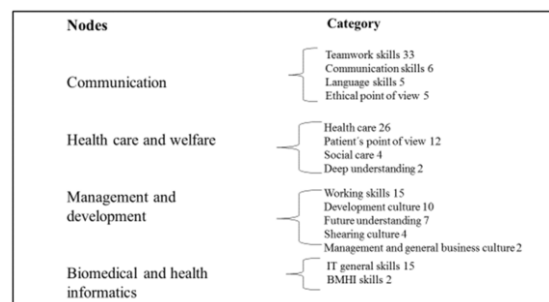


Figure 3 – Professional competences that the students from different branches bring to the multiprofessional group

Communication

The “Communication” node is the first node about the professional competence students from different branches bring to the multiprofessional group, and it was classified into three categories: 1) The “Teamwork skills” category has 33 items, including “*spreading ideas*” and “*the capability to work in a multiprofessional team*”; 2) The “Communication skills” category has six items, such as “*communication (written, spoken, tools, social media)*”; 3) The “Language skills” category has five items such as, such as “*good language skills*” and “*writing competence*”; and 4) The “Ethical point of view” category has five items, such as “*compliance*” and “*loyalty*”.

Health care and welfare

The second node is “Health care and welfare”, which was classified in four categories: 1) The “Health care category” has 26 items, including “*up-to-date information about public health care*” and “*working with chronic diseases and use of e-compliance*”; 2) The “Patient’s point of view” category has 12 items, such as “*a practical skillset and an understanding of how to prioritize the needs of a patient*” and “*inform, instruct and guide people in making the change to better in behavior and movement*”; 3) The “Social care” category has four items, for example, “*knowledge of social branch, law, practicals, and ethics*”; and 4) The “Deep understanding” category has two items, including “*a diverse work and internship experience, with an understanding of the complexity of the health care system*”.

Management and development

The third node is “Management and development”, which was classified into five categories. 1) “Management and general working skills” has 15 items, such as “*leadership skills*” and “*the capability to work in a multiprofessional team*”. 2) The “Development culture” category has 10 items, including “*professionalism in my own branch/sector*” and “*service development*”. 3) “Future understanding” has seven items, for example, “*dedication in seeking further knowledge to improve the system*” and “*multidimensional insights*”. 4) The “Shearing culture” category has four items, such as “*give and share my knowledge*”. Finally, 5) “Business culture” has two items, for example, “*knowledge in business, finance*”.

Biomedical and health informatics

The fourth node is “Biomedical and health informatics”, (BMHI) which was classified into two categories: 1) The “IT general skills” category has 15 items, such as “*IT background from end user support and IT management, customer service and support*”; and 2) “BMHI skills” has two items: “*knowledge concerning possibilities using digital services in public health care*” and “*specification and documentation skills*”.

Discussion

In the last decade, health and welfare services have become a customer-centric culture with many digital services [2]. Designing and innovating these new digital-service processes effectively require human-centered co-creation with sensitivity and motivation [3]. This study aimed to describe the competences that students had before studies and those they expect to gain from the study module “Developing Digital Health and Welfare Services” in multiprofessional groups as part of their bachelor studies.

The study results showed that the students expect to gain an understanding of the current situation in eHealth and social care services. They also want to acquire skills to communicate with customers through digital services and learn how to develop and

design services. Lifelong learning has become an even more important aspect because professionals and citizens need new competences to use eHealth services in flexible ways [4].

IMIA’s biomedical and health informatics curriculum [7], EDCL [6], and design-thinking competences [14] are the framework for the DeDiWe studies. Results showed that students are interested in understanding regulations and how quality customer-oriented services are built. They want to know the future trends in eHealth and welfare services, and also connect to different service providers and meaningful networks. Students prioritized the knowledge and skills needed to develop services in multiprofessional teams. They want concrete skills that will help them work with apps, but at the same time, do want to gain an understanding of development projects on an international level.

In developing digital health and welfare services in multiprofessional teams, the bachelor’s students see as equally important biomedical and health informatics knowledge and skills, a communicative culture, and high-quality services. The results of this study are in line with the EQF [13], which provides the framework for general professional competences to all professionals to reach the requirements to be active partners in research, development, and innovation processes. IMIA’s curriculum [7] is multiprofessional, between IT and health care, but business and social professionals are not included. Also for designers [15], BMHI studies could be useful. Today’s health care services need a paradigm change to have more open and customer-oriented services, in which all can take part in designing new services and implementing them actively in society [3]. In DeDiWe studies, we have had the opportunity to allow health and social care, IT engineer, IT business, and business bachelor’s students to take part in a common study module about developing digital health care services. This project is one intervention by the EU that supports enhancing the population’s digital skills and increasing the level of ICT professionalism [5].

The results of this study showed that students believe that they can bring their own competences to the multiprofessional team and they believe in their ability to co-create safe, customer-friendly digital health and welfare services for customers. These results differ from those of the EU’s [5] Single Market Strategy, which determined that nearly half of the EU population does not have the proper digital skills for jobs that require ICT competence, but these results are in line with World Health Organization’s [2] strategy’s objectives to have eHealth services implemented globally.

The questionnaire was developed for this study and contained only open-ended questions. The next step is to analyze the quantitative data from the questionnaire. The wording and format of the questions may have influenced the results. The data were analyzed using NVivo 10; however, because of a qualitative analysis, there is always a possibility of diverse interpretation among two researchers. The results are not generalizable because of the small sample size. The sample mainly reflects the opinions of the health care sector because most of the students represent that sector. However, these results imply that students are capable of working multiprofessionally in developing digital health and welfare services. Students were informed that completing the e-questionnaire was voluntary, but that we recommended it because of the importance of the project. This and the need for adequate English skills might have affected the response rate and the results of the study.

Conclusion

Successful development of multidisciplinary digital social and health care services requires changes and cooperation in education between various professions including health care, social care, IT, and business. Digital health and welfare services development need a multiprofessional curriculum and a combination of all teachers' competences for high-quality service development.

Acknowledgements

Funding: This work was supported by Interreg, Central Baltic, European Regional Development Fund as a part of the CB 25 Developer of Health and Welfare Service project.

References

- [1] S.T.M. (Ministry of Social Affairs and Health) 2015. Information Strategy for Social and Health Care 2020. Available at: http://www.julkari.fi/bitstream/handle/10024/125500/URN_ISBN_978-952-00-3548-8.pdf?sequence=1
- [2] World Health Organization, *Sixty-sixth World Health Assembly wha66.24*. Agenda item 17.5 27 May 2013 eHealth standardization and interoperability. Available at: http://apps.who.int/gb/ebwha/pdf_files/WHA66/A66_R24-en.pdf?ua=1
- [3] K. Freire and D. Sangiorgi, Service design & healthcare innovation: from consumption to co-production and co-creation. *Second Nordic Conference on Service Design and Service Innovation* (2010), 1-11. 1-3 December, Linköping, Sweden. Available at: <http://www.servdes.org/pdf/freire-sangiorgi.pdf>
- [4] Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning (2006/962/EC) Official Journal of the European Union. Brussels, 18 December 2006. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32006H0962>
- [5] Communication from commission to the European Parliament, the Council, the European economic and social committee and the committee of the regions. *A Digital Single Market Strategy for Europe* (2015). Brussels, May 6. 2015. COM (2015) 192 final.
- [6] European Computer Driving Licence Association, *ECDL and Qualifications Frameworks Worldwide* 2015. Available at <http://ecdli.org/policy-publications/ecdl-and-qualifications-frameworks-worldwide>
- [7] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, E. Hovenga, K.C. Lun, H. Marin, F. Martin-Sanchez, and G. Wright, *Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics*. *Methods Inf Med* 49, 2 (2010), 105-120. Available at: <http://dx.doi.org/10.3414/ME5119>
- [8] N. Staggars, C.A. Gassert, A. Carole, and C. Curran, Informatics competencies for nurses at four levels of practice. *J Nurs Educ* 40.7 (2001), 303-316.
- [9] N. Staggars, C.A. Gassert, and C. Curran, A Delphi study to determine informatics competencies for nurses at four levels of practice. *Nurs Res* 51.6 (2002), 383-90
- [10] B.W. Thompson and D.J. Skiba, Informatics in the nursing curriculum: a national survey of nursing informatics requirements in nursing curricula. *Nurs Educ Perspect* 29.5 (2008), 312-317.
- [11] TIGER Initiative, *Technology Informatics Guiding Education Reform* 2009: Informatics Competencies for Every Practicing Nurse: Recommendations from the TIGER Collaborative. Available at http://tigercompetencies.pbworks.com/f/TICC_Final.pdf
- [12] T. Naccarato, Child welfare informatics: A proposed subspeciality for social work. *Children and Youth Services Review*. 32, (2010), 1729-1734. doi:10.1016/j.childyouth.2010.07.016
- [13] The European qualifications framework for lifelong learning (EQF) European Commission. Education and Culture Lifelong Learning: Education and Training policies. Coordination of Lifelong Learning Policies. European Communities, 2008. Luxembourg. DOI 10.2766/14352 Available at: http://www.ecompetences.eu/site/objects/download/4550_EQFbroch2008en.pdf
- [14] Directive 2013/55/EU of the European parliament and of the council of 20 November 2013 amending directive 2005/36/EU on the recognition of professional qualifications and regulation (EU) no 1024/2012 on administrative cooperation through the internal market information system ('the imi regulation') Official Journal of the European Union. Brussels, 28 December 2013.134-173. Available at:

<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ.L.2013:354:FULL&from=EN>

- [15] R. Razzouk and V. Shute, What is design thinking and why is it important? *Rev Educ Res* 82 (2012), 330-348. doi: 10.3102/0034654312457429
- [16] S. Elo and H. Kyngäs. The qualitative content analysis process. *J Adv Nurs* 62.1 (2008), 107-115. doi: 10.1111/j.1365-2648.2007.04569.x

Address for correspondence

Outi Ahonen, Vanha maantie 9, 02650 Espoo, Finland. Email: outi.ahonen@laurea.fi, phone: +358400772053

A User-Centered Glucose-Insulin Data Display for the Inpatient Setting

Michael E. Bowen^{a,b}, Umme Rumana^a, Ethan A. Killgore^a, Yang Gong^a

^a School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Texas, USA,

^b The University of Texas Southwestern Medical Center, Dallas, Texas, USA.

Abstract

Inpatient management of insulin-dependent diabetes (IDD) is a complex task that requires clinicians to cognitively process information across distinct domains in different locations of the electronic medical record (EMR). Current data displays are not optimized to support insulin management by end users. We sought to develop a set of user-centered displays of capillary glucose data and insulin dose to improve inpatient management of IDD. Our proposed conceptual data display prototype is designed to simplify the presentation and visualization of key information needed for treatment decisions. The goal is also to enhance clinician's ability to identify opportunities to optimize insulin dosing and decrease end users' cognitive load and error rates.

Keywords:

Diabetes Mellitus; Blood Glucose; Data Display

Introduction

Glycemic control in hospitalized patients with type 2 diabetes is associated with improved outcomes [1]. Although inpatient diabetes management is a common task in clinical practice, glycemic control is often suboptimal. Physician management of inpatient diabetes involves three core tasks: a) reviewing Capillary Blood Glucose (CBG) data; b) interpreting values in relation to insulin dosing and meal intake, and c) determining if a change in treatment plan is necessary.

Unfortunately, data needed to support insulin adjustments are sub-optimally organized and displayed in the electronic medical record. For example, capillary glucose measurements are reported in the laboratory results, insulin dosing and administration are presented in the medication administration record, and the relationship of glucose readings to meal intake is derived from electronic time-stamps on the glucose readings as they relate to insulin orders. This requires processing and integration of data from multiple locations, multitasking, and attention switching [2]. Given the limited memory span of 7 ± 2 [3], mental juggling between tasks can diminish information processing speed, resulting in inefficiencies in completing a common, routine clinical task, duplication of effort, physician burnout, and potentially suboptimal diabetes management in the inpatient setting.

Novel approaches are needed for representing and consolidating the inpatient diabetes management data into intelligent displays. In this paper, we focused on the treatment of hospitalized patients treated with basal-bolus insulin who undergo routine CBG monitoring before breakfast/lunch/dinner and at bedtime. Our proposed display will assist clinicians with the review and interpretation of CBG values by merging data (CBG results, CBG time, and insulin administration) from multiple EMR dimensions into a single location.

Background

Iatrogenic, unintentional Adverse Drug Events (ADEs) are among the most common causes of inpatient complications in the United States [4]. Hospital-acquired ADEs extend the length of stay, result in higher healthcare costs, and may cause tragic loss of life [5-6]. Injected medications, including insulin, account for 63% of ADEs and about 50% are believed to be preventable [7]. Injectable drugs are associated with an estimated 1.2 million preventable inpatient ADEs resulting in an average annual increase of \$600,000 per hospital [8]. There is great concern about the high risk of preventable hypoglycemia related to insulin use [8]. Integrated data displays are needed to improve inpatient glucose control, improve safety, and increase clinician efficiency.

Fonda et al. [9], conducted a focus group study with diabetes patients to understand patients' needs and tasks to inform the design and functionality of a diabetes personal health application. Although this study focused on outpatient diabetes self-management by patients, not providers, core principles from this study were used to inform the development of our inpatient display. In this study, patients reported that the collection, recording, and collating of glucose data for diabetes management is a burdensome process. To help address this, researchers highlighted the need for actionable, understandable, timely information to support glucose management – especially as it relates to interactions with medications, diet, and lifestyle. Fonda et al. [9] utilized flexible and reusable gadgets built on web-based applications within the data display portal, to integrate data from different domains into a consolidated display. For example, the “glucose gadget” displays data from glucose monitors, the electronic medical record, or patient-entered data on a temporal graph. The extensible “gadget” concept suggested our proposed dimensional flexibility. Kumar et al. [10] developed an approach to utilize remotely collected data from continuous glucose monitors, import, and store them into EMR flowsheets, and interface with a novel data display (GluVue). In feasibility testing, the GluVue efficiently collated large volumes of data and improved clinicians' and patients' ability to identify opportunities to improve glucose management [11]. It was this concept of novel display of outpatient glucose data that further informed our ideas relating to potential applicability to management of inpatients as well.

Theoretical Framework and Methods

Our proposed data display was based on three visualization theories—distributed cognitive theory, relational data display framework, and Gestalt's principles. Glucose-insulin visuals were created using Tableau® 9.3. The following sections will describe theories and their applications in our display.

Distributed Cognitive Theory

When clinicians interact with computers, their cognition is distributed in two spaces—internal and external. The internal space represents clinician factors such as attention, memory, knowledge, experience, language, reasoning, and decision-making. The external space represents computer factors, such as user interface design, screen navigation, and data exhibition and organization, among others [12]. For improved decision-making, we aimed to reduce distributed cognitive load related to short-term memory and data display interface.

Table 1- Description of CBG Values by Date, Finger-Stick Time, and Finger-Stick Event

Date	Day	Finger-Stick Time (24 hr)	Finger-Stick Time (Event)	CBG Levels (mg/dl)
9/18/2016	Day 1	6:00	Pre-Breakfast	82
9/18/2016	Day 1	12:00	Pre-Lunch	172
9/18/2016	Day 1	18:00	Pre-Dinner	191
9/18/2016	Day 1	22:00	Bedtime	220
9/19/2016	Day 2	6:00	Pre-Breakfast	118
9/19/2016	Day 2	12:00	Pre-Lunch	160
9/19/2016	Day 2	18:00	Pre-Dinner	175
9/19/2016	Day 2	22:00	Bedtime	190
9/20/2016	Day 3	6:00	Pre-Breakfast	94
9/20/2016	Day 3	12:00	Pre-Lunch	185
9/20/2016	Day 3	18:00	Pre-Dinner	167
9/20/2016	Day 3	22:00	Bedtime	191

Table 2- Description of Insulin Dose by Date, Day, Injection Time, and Subcutaneous Insulin Dose

Date	Day	Injection Time (24 hr)	Injection Time (Event)	Insulin Dose (Units/d)
9/18/2016	Day 1	6:00	Pre-Breakfast	6
9/18/2016	Day 1	12:00	Pre-Lunch	12
9/18/2016	Day 1	18:00	Pre-Dinner	14
9/18/2016	Day 1	22:00	Bedtime	30
9/19/2016	Day 2	6:00	Pre-Breakfast	6
9/19/2016	Day 2	12:00	Pre-Lunch	12
9/19/2016	Day 2	18:00	Pre-Dinner	14
9/19/2016	Day 2	22:00	Bedtime	30
9/20/2016	Day 3	6:00	Pre-Breakfast	6
9/20/2016	Day 3	12:00	Pre-Lunch	12
9/20/2016	Day 3	18:00	Pre-Dinner	14
9/20/2016	Day 3	22:00	Bedtime	30

Clinicians process external, visual information by recognition and recall. Our informal task analysis done for design purposes by the authors, from the perspective of a physician's task of inpatient management of insulin dosing of patients with Type II diabetes indicates that current physician-EMR interactions and information processing flow require 13 steps to make decisions and are discussed as follows. Our intention is to use this analysis to define the framework that informs our proposed design, which will undergo evaluation during future iterations.

1. The physician's undivided *attention* is needed to process information and navigate the EMR.
2. The physician *retrieves* a laboratory screen to view the results; the lab screen icon/menu corresponds to an *external* representation.
3. S/he *consults* the tabular display on Table 1 to read the CBG level; the tabular layout constitutes an *external* representation.
4. S/he *perceives* CBG levels as high/normal/low, referred as pattern recognition, which is an example of an *internal* representation.
5. Prior knowledge stored in *long-term memory* is then applied to understand CBG patterns, diagnostic, and

treatment criteria, all of which are derived from an *internal* representation. Abnormal CBG values may necessitate insulin adjustment.

6. The physician then assigns abnormal CBG values to *short-term memory*.
7. S/he *switches* task and attention to locate medication screen in the EMR.
8. S/he *retrieves* medication screen located on a separate tab in the EMR.
9. S/he *consults* the tabular display on Table 2 to read the insulin dose.
10. S/he *perceives* insulin levels as high/appropriate/low, referred as pattern recognition.
11. Prior knowledge stored in *long-term memory* is then applied to understand insulin dosage, diagnostic, and treatment criteria.
12. Again, knowledge is combined with the *recalled* CBG level, and a decision is made about a dose change.
13. If adjustment is needed, medication *order entry* screen is consulted, if not, monitoring continues.

The entire process is delayed since two screens have to be consulted with several numbers in mind. Repetition of the process for different patients results in the unnecessary investment of time, click-burden, and stress.

To reduce mental strain related to the recall, we eliminated the internal/intrinsic tasks linked to short-term memory by reducing the external/extrinsic load. A single, integrated screen model is created by merging two relational tabular displays that will limit the decision-making steps to 6, described as follows:

1. The physician's undivided *attention* is needed to process information and navigate the EMR.
2. The physician *retrieves* an integrated data display screen to view glucose-insulin values.
3. S/he *consults* the new graphical displays that are introduced in the results section below.
4. S/he *perceives* CBG levels as high/normal/low, referred as pattern recognition in context with high/appropriate/low insulin levels.
5. Prior knowledge stored in *long-term memory* is then applied to understand CBG-insulin patterns, diagnostic, and treatment criteria. Abnormal CBG values may necessitate insulin adjustment.
6. If adjustment is need, medication *order entry* screen is consulted, if not, monitoring continues.

Relational Data Display Framework

A common theme across papers surveyed, beginning with Hutchins et al. [13] in 1985, and as recently as Gong et al. [14] in 2009, is that the closer the internal representations of a task are to the external representations depicted through the dimensions of a relational display, the more successfully a clinician user can complete a search task and reach their goal [13-15]. More specifically, the proximity of scale between the represented and representing dimensions, particularly when numerical clinical data are involved, is key in providing the optimal displays [14]. This is also implicitly noted in the work of Bauer and colleagues [16] in that maximizing the degree of external representations, both as tables and graphs, can improve clinician understanding and subsequent action, though the representational effect is always a factor in creating a disparity between clinicians' understanding of the data.

Table 3- Dimensional Properties and Rationale

Dimensions	Representational Dimensional Scale	Representing Dimensional Scale	Rationale
Day	Interval	Interval	The CBG values recorded each day are associated with a date that corresponds to the requirements for interval scale in the visualization.
Time	Interval	Ordinal	Points in time itself are interval in scale, however, each of the four-time points in the visualization have only category and magnitude, but do not occur at equal intervals, nor is there an absolute zero value possible since we are still only observing points in time.
CBG	Ratio	Ratio	The values for CBG are all categorical, have magnitude, equal interval, and could potentially have a value of absolute zero.
Insulin	Ratio	Ratio	The values for insulin are all categorical, have magnitude, equal interval, and could potentially have a value of absolute zero.

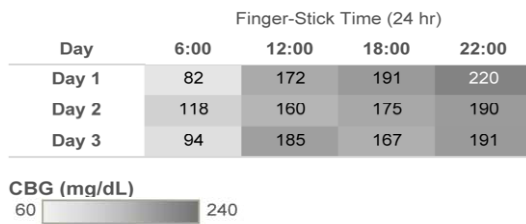


Figure 1- Highlight table: CBG Broken Down by Finger-Stick Time vs. Day

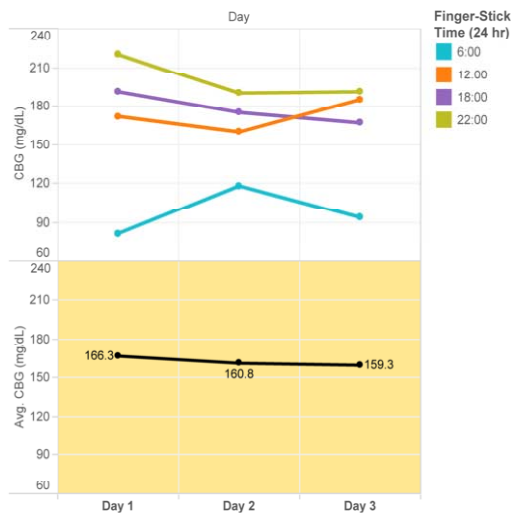


Figure 2- Line chart: the Trends of CBG and Average of CBG for Day

Properties of the display are determined by properties of its dimensions. Dimensional properties are measured in scales. The four scales types include nominal, ordinal, interval, and ratio. Nominal scale is the lowest form of measurement scale. The numeric values classify characteristics of people, objects, or events into categories. Ordinal scales have categories arranged in some meaningful order without magnitude. In interval scales, the distance between ordered category values are equal. Ratio scales are most precise level of measurement

and possess magnitude, equal intervals, and an absolute zero. The scales selected for our project are given in Table 3.

Gestalt's Principles of Visual Perception

Relational and/or graphical displays that optimally align the cognition and perceptions of innately distinct objects into related groups can minimize clinicians' cognitive load while maximizing memory and recall [17]. Four Gestalt's principles enhance cognition and improve decisions including similarity, proximity, connection, and enclosure.

Method of Evaluation for Design Purposes

An initial inspection was conducted by the authors and then an extended to group of 13 with extensive experience in the clinical informatics domain. The inspection was informal in nature, with the intention of soliciting empirical responses to guide further development.

Results

The goal of the highlighted table in Figure 1 is to facilitate pattern recognition of high and low values. The display is grouped based on Gestalt' similarity principle. The grey CBG gradient filter is grouped based on similarity and proximity to display the glycemic values. Color distinguishes abnormal values for selective attention and hastening information processing, aligning with distributed cognition principles. We also introduced the dimension of time as an ordinal scale, according to relational data display principles. The goal of the line chart in Figure 2 is to display longitudinal CBG values stratified by time of day on the top panel and average daily glucose value on the bottom panel. This display improves upon current glucose displays in the EMR which are most commonly time-ordered tabular displays that do not clearly group values at similar time points (i.e. breakfast values at 6 am are not grouped together but time-ordered between 22:00 values and 12:00 values). Consistent color patterns are chosen to highlight correlated data across the graphs (i.e. all 06:00 glucose values are blue). The display is grouped based on Gestalt's law of similarity, proximity, connection, and enclosure, as well as building on the distributed cognition principle of proximity compatibility.

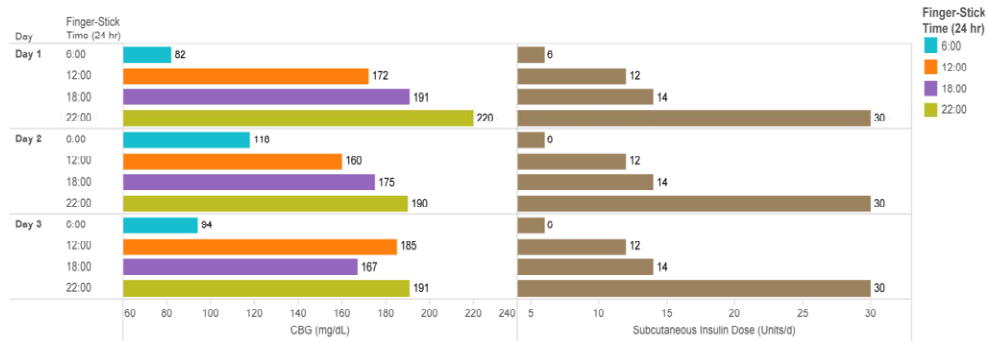


Figure 3- Horizontal Bar Chart: CBG and Subcutaneous Insulin Dose for Each Finger-Stick Time Broken Down by Dday

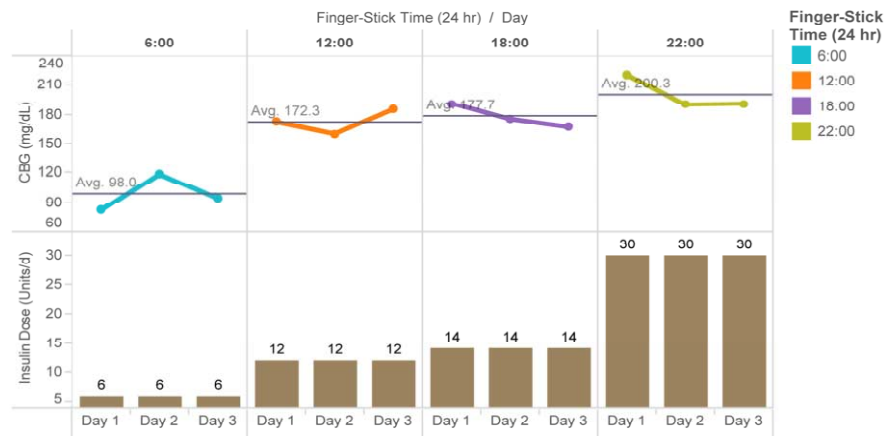


Figure 4- Line Chart-Vertical Bar Chart: the Trends of CBG and Subcutaneous Insulin Dose for Day Broken Down by Finger-Stick Time

The goal of horizontal bar chart in *Figure 3* is to display both the capillary glucose reading and the corresponding insulin dose administered at the concurrent time point. This allows clinicians to correlate the insulin dose with the corresponding glucose reading to assess the appropriateness of the dosing. For example, the 30 units of insulin administered at 22:00 achieves adequate control at 06:00 (glucose range 82-118 mg/dL); however, the 14 units administered at 18:00 with dinner does not match the carbohydrate intake at dinner, as reflected by the elevated readings at 22:00 (range 190-220 mg/dL). The display is grouped using Gestalt's similarity, proximity, and enclosure principles. Distributed internal and cognitive load is reduced by merging two disparately located data tables on a single screen.

The goal of the line chart-vertical bar chart in *Figure 4* is to displays the average glucose for each time period and the insulin dose administered at that time. This display is particularly important because clinicians often want to trend summary glucose data over several days prior to adjusting insulin dosing. This helps avoid over or under reacting to any single value. However, it still allows for easy recognition of hypoglycemia, an important safety concern, along with the dose of insulin administered that may be potentially related to the low glucose value. The display is grouped based on Gestalt's law of grouping and the distributed cognitive load is lowered by merging table 1 & 2. The inherent relational data display principles are also applied in further representation of internal dimensions as external dimensions depicted in such a way as to reduce the difficulty of clinician understanding and appropriate intervention.

From the initial inspections by the authors, as well as the empirical evaluations of the extended informatics group, the set of theory-grounded data display was well received. Figure 4 below was especially commented as the easiest to interpret, because "having insulin doses and glucose values displayed in proximity reduced the need for screen switching and internal representation, appreciated by clinicians."

Discussion

Optimized management of patients with insulin dependent diabetes in the hospital setting is critical to improved patient outcomes since such patients have increased morbidity, mortality, and longer lengths of stay. Clinicians must monitor patients' capillary blood glucose level and insulin dosage, which necessitates information from different parts of an EMR system to analyze and intervene cognitively. Today's prevailing EMR data displays are inadequate to meet clinicians' cognitive needs, due to data storage in different places within the system that interrupt cognition or presented in tabular formats that make it challenging to identify actionable trends. This also increases cognitive load, reduces intellectual performance, increases errors, and ultimately results in sub-optimal task completion [18]. Our belief is that if related variables required for management of diabetic patients are well represented externally via a relational data display harnessing principles of distributed cognition, then the clinician tasked with managing glycemic management will have less cognitive load and fewer potential interruptions from

information foraging, which may lower or eliminate associated cognitive costs.

Limitations and Future Work

Although the proposed display is grounded in theory, the development team has the complementary expertise in clinical, informatics and programming skills. It is essential to conduct a formal evaluation regarding its effectiveness and efficiency in clinical practice. Future work including a fuller analysis of task, users, representation, and function is necessary to inform experimental design as proposed by Zhang et al. [19]. This will provide further user feedback prior to formal experimental testing. Additional work will include options to change the representing dimension for time from an ordinal to an interval scale to expand the evaluation of variables across a wider time span.

To determine the proposed display's effectiveness in clinical practice and potential to decrease cognitive load, it will need to be integrated into an EMR. Only then can the implementation, usability, and effectiveness be evaluated. Study results and user feedback will inform ongoing opportunities for improvements in the data display and identify additional clinical variables amenable to similar relational displays.

Conclusion

Overall, our proposed insulin-glucose data display strategy seeks to: reduce the intrinsic load by minimizing the screen navigation task and clicks, reduce the extrinsic/external load by merging insulin dosage and glucose level tables into one display, and reduce the extrinsic and germane load. Theory-based relational information displays using principles of cognition, distributed cognitive tasks, and Gestalt theory can help address challenges in inpatient glucose management. All told, design collaborations between informaticians, computer programmers, and clinicians are essential to solving common clinical problems.

Acknowledgements

Dr. Bowen was supported by the National Institute of Diabetes, Digestive, and Kidney Diseases (NIH/NIDDK K23 DK104065). Dr. Gong was in part supported by The Agency for Healthcare Research & Quality (1R01HS022895), and University of Texas System Grants Program (#156374).

References

- [1] G. E. Umpierrez, R. Hellman, M. T. Korytkowski et al., Management of hyperglycemia in hospitalized patients in non-critical care setting: an endocrine society clinical practice guideline, *J Clin Endocrinol & Met* 97 (2012), 16-38.
- [2] D. Fougner, The relationship between attention and working memory, In N. B. Johansen (Ed.), *New research on short-term memory*, New York, NY: Nova Biomedical Books, (2008), 1-45.
- [3] G.A. Miller, The Magical Number Seven, Plus Or Minus Two: Some Limits on Our Capacity for Processing Information, *The Psychological Review* 63 (1956), 81-97.
- [4] A. Sawarkar, C. A. Keohane, S. Maviglia et al., Adverse Drug Events caused by Serious Medication Administration Errors, *BMJ Quality & Safety* 21 (2012), 933-938.
- [5] D.C. Classen, S.L. Pestotnik, R. Evans, et al., Adverse drug events in hospitalized patients: Excess length of stay, extra costs, and attributable mortality, *Journal of American Medical Association* 277 (1997), 301-306.
- [6] A.J. Weiss, A. Elixhauser, J. Bae, and W. Encinosa, Origin of Adverse Drug Events in U.S. Hospitals, 2011. The HCUP report: Healthcare cost and utilization project (HCUP): *Statistical Briefs* 158, Agency for Healthcare Research and Quality, Rockville, MD, 2013.
- [7] Adverse events in hospitals: national incidence among Medicare beneficiaries, Washington, D.C.: Department of Health and Human Services, Office of Inspector General, 2010.
- [8] B. J. Lahue, B. Pyenson, K. Iwasaki, et al., National Burden of Preventable Adverse Drug Events Associated with Inpatient Injectable Medications: Healthcare and Medical Professional Liability Costs, *American Health & Drug Benefits* 5 (2012), 1-10.
- [9] S. J. Fonda, R. J. Kedziora, R. A. Vigersky, and S. E. Bursell, Combining iGoogle and personal health records to create a prototype personal health application for diabetes self-management, *Telemedicine and e-Health* 16 (2010), 480-489.
- [10] R. B. Kumar, N. D. Goren, D. E. Stark et al., Automated integration of continuous glucose monitor data in the electronic health record using consumer technology, *Journal of the American Medical Informatics Association* 0 (2016) 1-6.
- [11] GluVue: A real-time clinical decision support tool for interpreting blood glucose data, 2016.
- [12] J. Zhang, and D. A. Norman, Representations in distributed cognitive tasks, *Cognitive Science* 18 (1994), 87-122.
- [13] E. Hutchins, J. Hollan, and D. Norman, Direct manipulation interfaces, *Human-Computer Interaction* 1 (1985), 311-338.
- [14] Y. Gong, and J. Zhang, Toward A human-centered hyperlipidemia management system: The interaction between internal and external information on relational data search, *Journal of Medical Systems* 35 (2011), 169-177.
- [15] D. T. Bauer, S. Guerlain, and P. J. Brown, The design and evaluation of a graphical display for laboratory data, *Journal of the American Medical Informatics Association* 17 (2010), 416-424.
- [16] J. Zhang, A representational analysis of relational information displays, *International Journal of Human - Computer Studies* 45 (1996), 59-74.
- [17] S.E. Palmer, Vision Science: Photons to Phenomenology, Massachusetts Institute of Technology Press, Cambridge, Massachusetts, USA, 1999.
- [18] R. E. Mayer, and R. Moreno, Nine ways to reduce cognitive load in multimedia learning, Educational psychologist, Taylor & Francis, 2003.
- [19] J. Zhang and M. F. Walji, M. F., TURF: Toward a unified framework of EHR usability, *Journal of Biomedical Informatics* 44 (2011), 1056-1067.

Address for correspondence

Yang Gong, MD, PhD
7000 Fannin St, Suite 600,
Houston, Texas, USA
Email: Yang.Gong@uth.tmc.edu
Tel: +1-713-500-3547

Lack of Training for a Downtime Procedure: End User's Perceptions of an Electronic Health Record Contingency Plan

Rodrigo Cano, Brenda Bejarano, Mercedes Vidal, Daniel Luna, Sonia Benítez

Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

The implementation of an Electronic Health Record has many benefits; but when it is not available, it can impact patient continuity of care. If there is no support, or a failure to guarantee the continuity of services, contingency plans have to be implemented to overcome the information disruption. End users are in direct contact with the information system, and are responsible for documenting patient clinical information. Focusing on them, we propose the design, development, and validation of a survey to evaluate the beliefs, knowledge, and perceptions of end users, about the Electronic Health Record contingency plan. Preliminary results showed that even when there were less downtime periods, end users perceived that they did not have adequate training or information about how to go through the downtime event.

Keywords:

Electronic Health Records; Perceptions; Health Care Surveys

Introduction

Following the impulse given by different Electronic Health Records (EHR) Incentive Programs, the adoption of a Health Information System by organizations has increased [1]. Many benefits are attributable to EHR adoption [2,3], but when it is not available, it can impact health services costs and the patient continuity of care [4]. If problems related to standardization and protection of critical operative processes and relevant information are not solved, it is not possible to maintain the continuity of care [5]. The nature of hospital procedures has an impact over the patient's information management. If there is no support, or a failure to guarantee the continuity of services, different procedures should be initiated to protect important information [6,7]. Institutions must implement contingency plans for the maintenance of their systems during critical moments, and to overcome the information disruption. There are guides, such as SAFER, that assesses socio-technical aspects of contingency plans and identifies safety practices associated with EHR unavailability [8].

In that context an important point is the way that end users are engaged. In general, contingency plans tend to prioritize technical aspects [9,10] and responsiveness of it [11]. The importance of taking into account end users is because they are in direct contact with the information system, and are responsible of documenting the patient clinical information. Focusing on end users, we proposed the design, development and validation of a survey to evaluate the beliefs, knowledge, and perceptions from end users, about the EHR contingency plan.

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA) is a non-profit health care academic center, with over 2,800 physicians, 2,800 other health team members (including 1,200 nurses) and 1,900 administrative and support employees. HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 1,000 home care beds, 25 outpatient clinics, and 250 associated private practices located in Buenos Aires city and its suburban area. It has a Health Maintenance Organization that covers more than 150,000 people and also provides health services to other people who are covered by affiliated insurers. In addition, HIBA is a teaching hospital, with over 30 medical residency-training programs and 34 fellowship programs [12].

Survey

This is a mixed method study that included a survey design and validation. We designed the survey in 4 phases:

- Phase 1: Literature review and qualitative research
- Phase 2: Panel Expert Review
- Phase 3: Qualitative pretest
- Phase 4: Survey Validation

Phases 1, 2 and 3 are described in [13]. We continued analyzing phase 3 in this paper, recruiting new participants. In preliminary phases, the survey was paper based, in successive iterations we used Google® Forms, and finally we implemented the survey using REDcap® platform. We recruited participants through snowball sampling and key users representatives. The validation phase was carried out on active personnel of HIBA in 2015. They performed the following tasks using the EHR:

- Data visualization
- Image and lab results visualization
- Clinical notes
- Vital signs
- Request of laboratory orders
- Drugs prescriptions
- Referral Request
- Drug administration documentation

Exclusion criteria were:

- Refusal to participate in the study
- Workers with less than 6 months in the institution
- EHR super-users

- Personnel that have not used the EHR in the last 12 months

We used Stata® 13 and Microsoft Excel® for the quantitative and descriptive analysis. For validation and to evaluate the internal consistency for complete sections of the survey, we used Cronbach's alpha. In addition, we included semi-structured interviews to respondents, looking for agreements, disagreements, and new domains.

This research did not imply any additional risks for participants, and was executed in accordance with national and international norms. All data of the study was treated with maximum confidentiality and in anonymous way.

Results

During the **pre-testing phase** we evaluated and assessed the post-contingency of a planned downtime. Seventeen participants that completed the survey were interviewed with a semi-structured questionnaire, and they commented aspects related to wording and semantics. The identified domains with previous phases were:

- "If there is a downtime of the system or it is unavailable"
- "About the communication of the downtime"
- "During the downtime or unavailability"
- "When the system is available again"
- "Considering the downtimes of the EHR over time"

The most important findings were related to "communication", "an alternative work plan", and "training". This resulted in the creation of new variables, which were: knowledge of the procedure, plan implementation, end user's response capacity, accessibility during downtime, post-contingency data restoration, communication between the health informatics department and other areas, workflow impact, registration possibilities and time needed to activate the plan. The result was the *pre-tested qualitative survey*. For this phase, participants' characteristics can be seen in Table 1.

Table 1 – Pre-test phase participants' characteristics

Characteristics	Participants n=17
Female (%)	41
Age Range(years)	21-51
Occupation	
Physicians (%)	71
Nurses (%)	24
Administrative personnel (%)	6

The number of resident physicians and attending physicians were balanced. 35% of physicians have an internal medicine certification, 12% were endocrinologists, 6 % pulmonologists, 6 % pediatricians, and 6 % were orthopedists, among others. About the working place, 59% in inpatient care, 35% worked in ambulatory care, and 6% in emergency care. The length of service was between 1 to 5 years for 35%, 6 to 15 years for 24%, and more than 16 years for 35% of the participants.

The **validation phase** started in the second semester of 2015, and it is planned to finish in the first semester of 2017. We sent a new version of the survey, and invited the participants to a semi-structured interview when the survey was completed. Characteristics are shown in Table 2. The result of this process was a *validated survey*. About the completeness of the survey, thirty seven (72%) did fully completed it and thirty three (64%) of the participants did not perform

supervision or coordination tasks. Twenty two (43%) participants had an alternative plan for downtimes and thirteen mentioned that it was not documented. Thirty (58%) participants denoted not haven received training about the actions that have to be performed during a downtime, and twenty four (80%) answered that they would like to receive appropriate training. About the need to communicate with the Health Informatics Department, thirty three (64%) said that it was not necessary to establish communication, and seventeen stated that they needed to communicate with another area of the hospital. Twelve participants talked about not knowing where to obtain information about contingency plans or downtime actions to carry on, but 83% of them indicated that it would be useful. Only two of the participants who knew the contingency plan, answered that the web page with information about contingency was useful.

Table 2 – Validation phase participants' characteristics

Characteristics	Participants (n = 51)
Female (%)	63
Interviewed; n (%)	27(53)
Occupation	
Physicians (%)	28
Nurses (%)	21
Other (%)	10
Health technician (%)	6
Administrative personnel (%)	35

About the working place, 56% worked in ambulatory care, 39% in inpatient care, and 5% in emergency care. The length of service was between 1 to 5 years for 35%, 6 to 15 years for 22% participants, and more than 16 years for 27% of the participants. Related to the time needed in working hours to communicate a downtime, 43% of them mentioned "less than 1 hour", 14% more than 1 hour, and 43% did not know and could not estimate the time.

In relation to the access of information needed to work from the EHR, 85% stated that they could not access needed information. 80% mentioned that during downtimes they could not record clinical or other information required, in the patient EHR. 59% considered the information documentation in the EHR was affected by downtimes.

About perceptions of the frequency of downtimes along time, four (10%) participants considered that there are more downtimes than before, twenty five (64%) that there are less downtimes, six (15%) referred that there are the same events than before, and four (10%) did not give an answer.

The estimated time to complete the survey was around 5 minutes, and the estimated time to complete the interview was 15 minutes. The calculated Cronbach's alpha for structured and completed items from the domains identified was 0.76.

In addition we compared results to assess agreement between the presented answers in the interview compared to the survey answers. The qualitative analysis of the interviews is still in progress, up to now we did not find any major discrepancies with the results of the survey.

Discussion

The unavailability of the EHR during downtimes could result in important patient safety concerns in institutions that are not prepared to confront the contingency [8], for this reason we designed a survey to evaluate beliefs, knowledge, and perceptions that end users have, in relation to the contingency

plan of the EHR. We identified formerly ignored domains, and even with partial results of the preceding survey versions, this reassures the need of taking into account end users to design contingency plans.

In this phase the participants were from non-critical areas, with proper experience in the hospital (more than 1 year of length of work in their areas), but just half of them knew the exact procedures to apply contingency protocols, with previously described consecutive problems related to loss of information and patients' care [14]. Likewise, we found that the majority of participants referred that they wanted to establish a work plan, either protocolized, or created according to prior experiences of their areas. In addition, most participants reported not have received training on the contingency work plan during downtimes. Although there is a work plan created in case of downtimes, the personnel find their own contingency mechanisms, which responded to the needs created among working areas, which according to participants, have been effective and sustainable over time. It is also worth mentioning the great acceptability that they have to receive training in the area of contingency.

Another interesting point is related to the communication needs between areas during a downtime, since there is evidence that communicating planned downtime in advance will minimize frustration and confusion about why the system is unavailable [15], we found that a minority had the need to communicate with the Health Informatics Department. We consider that this is due to existence of mechanisms already created in the different areas, where they can overcome the eventuality without requiring external advice. It is interesting to mention that most respondents described needing a relatively short time for downtime warnings. This may be a reflection of how organized they are with their own contingency mechanisms, to perform actions such as patient support, medical appointment assignments, and general consultations. However, in critical areas, we noted that they prefer being notified earlier in order to coordinate complex activities to preserve a good performance of their activities. As it is usual to expect during downtimes, most respondents informed not being able to access clinical records and information that they required. This could also mean a loss of data that would be impossible to recover later, with possible adverse consequences in the patient's future care. Despite this, most of the respondents mentioned that there are fewer downtimes events in relation to former years, and thus greater stability and confidence in the system and this is concordant with preceding findings that shows the expression of the stability that end users tend to demonstrate in association to periodical revision of contingency plans [16].

As a limitation of the work we find that there is certain resistance to answering the survey, apparently not because of the particular topic of the survey, but to find time during work hours to answer the questionnaire. We did not consider that there is selection bias, having recruited representative samples of different working areas of the hospital. The potential information bias that could be found in surveys was improved during each phase creation and process, with the design and identification of different domains [17]. As strengths of the survey we found that it was possible to design it to be completed in an average of five minutes. Because this is a cross sectional single-center study we cannot generalize our findings to be valid to assess similar dimensions in end users of others health centers. Future multi-centered studies are needed to assess the generalizability of our results.

The internal consistency was acceptable, but only reliability does not imply validity. The validation is a process that started in the first publication that we made before [13]: the first step

was to establish face validity, and then the content validity with the Panel Expert Review; the next step was pre-testing the survey on a subset of our intended population, and finally was important to check the internal consistency of questions. We presented in this paper the pre-test and the evaluation of internal consistency. We are not using any score; therefore we did not perform any test-retest reliability. There were no possibility to evaluate convergent or discriminant validity or criterion validity because there is no gold or reference standard to compare. Iterative work is needed to develop future implementation for different institution's structures.

The partial qualitative analysis of interviews did not show any major discrepancies at the moment. Once we finish the qualitative analysis and the survey is implemented, it could be used to assess different contingency plans in any institution that uses an EHR, as the identified domains are common in health environments related to patient care. The assessment of the different domains and subsequent knowledge could lead to new interventions that could provide adequate training to face planned or unplanned downtime events. Related to this we are planning to do simulations for training and evaluation of knowledge of contingency procedures by the second semester of 2017.

Conclusion

This paper showed the continuation of our study about perceptions, beliefs, and knowledge of end users related to the contingency plan of EHR. This is an important topic to explore, that could lead to potential actions to improve mechanisms related to proper design and execution of future contingency plans. In our case, end users perceived that they did not have adequate training or information about how to go through the downtime event. Even with less downtime periods, the end users perceived the need of training and communication.

Acknowledgements

Funding was provided by Hospital Italiano de Buenos Aires.

Authors did not report any conflicts of interest. We thank all of the participants and collaborators especially Marcela T. Fernandez.

References

- [1] Electronic Health Records (EHR) Incentive Programs. 2016. Available from: <https://www.cms.gov/Regulations-andGuidance/Legislation/EHRIncentivePrograms/index.html?redirect=/EHRIncentiveprograms/>
- [2] The Benefits of Electronic Health Records (EHRs) | Providers & Professionals | HealthIT.gov . Available from: <https://www.healthit.gov/providers-professionals/benefits-electronic-health-records-ehrs>
- [3] N. Menachemi and T.H. Collum, Benefits and drawbacks of electronic health record systems, *Risk Manag Healthc Policy* 4 (2011), 47-55. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22312227>
- [4] P. Hartzband and J. Groopman, Off the record – avoiding the pitfalls of going electronic, *N Engl J Med* 358(16) (2008), 1656-1658. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18420497>
- [5] J. Gervas, Longitudinalidad y continuidad de la atención al paciente y a la familia: ayuda tecnológica a un problema complejo sin soluciones sencillas.
- [6] E. M. Campbell, D.F. Sittig, K. P. Guappone, R.H. Dykstra, and J.S. Ash, Overdependence on technology: an unintended adverse consequence of computerized provider order entry. *AMIA . Annu Symp proceedings AMIA Symp.* 2007; 94-98. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18693805>
- [7] J.K. Liker, C.J. Haddad, and J. Karlin, Perspectives on technology and work organization, *Annu Rev Sociol* (25) (1999), 575-596.

- [8] D. Sittig, Contingency Planning General Instructions for the SAFER Self Assessment Guides, 2014.
- [9] S. Menon, H. Singh, A.N.D. Meyer, E. Belmont, and D.F. Sittig, Electronic health record-related safety concerns: a cross-sectional survey, *J Healthc Risk Manag* 34(1) (2014), 14-26. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25070253>
- [10] J. Lei, P. Guan, K. Gao, X. Lu, Y. Chen, Y. Li, Q. Meng, J. Zhang, D.F. Sittig, and K. Zheng, Characteristics of health IT outage and suggested risk management strategies: an analysis of historical incident reports in China, *Int J Med Inform* 83(2) (2014), 122-130. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24246272>
- [11] M. Khalifa, Evaluating nurses acceptance of hospital information systems: a case study of a tertiary care hospital, *Stud Health Technol Inform* 225 (2016), 78-82. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27332166>
- [12] F. Plazzotta, D. Luna, and F. González Bernaldo de Quirós, Health information systems: integrating clinical data in different scenarios and users, *Rev Peru Med Exp Salud Publica* 32(2) (2015), 343-351. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26338397>
- [13] M.T. Fernández, A.R. Gómez, A.M. Santojanni, A.H. Cancio, D.R. Luna, and S.E. Benítez, Electronic health record system contingency plan coordination: a strategy for continuity of care considering users' needs, *Stud Health Technol Inform* 216 (2015), 472-476. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26262095>
- [14] M.I. Harrison, R. Koppel, and S. Bar-Lev, Unintended consequences of information technologies in health care--an interactive sociotechnical analysis, *JAMA* 14(5) (2007), 542-549. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17600093>
- [15] C.G. Fahrenholz, L.J. Smith, K. Tucker, and D. Warner, Plan B: a practical approach to downtime planning in medical practices, *J AHIMA* 80(11) (2016), 34-38. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19953791>
- [16] S. Ramthun, E.M. James, N. Hangiandreou, and C.E. Bender, Electronic imaging system implementation at Mayo Clinic Rochester: downtime procedures and communication plans, *J Digit Imaging* 13(2 Suppl 1) (2000), 206-207. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10847403>
- [17] J.S. Armstrong and T.S. Overton, Estimating nonresponse bias in mail surveys, *Publ J Mark Res* 14(3) (1977), 396-402. Available from: http://repository.upenn.edu/marketing_papers

Address for correspondence

Sonia Benítez

sonia.benitez@hospitalitaliano.org.ar

Chile's National Center for Health Information Systems: A Public-Private Partnership to Foster Health Care Information Interoperability

^aDaniel Capurro, ^bAisen Echeverry, ^cRosa Figueroa, ^dSergio Guíñez, ^eCarla Taramasco,
^fCésar Galindo, ^gAngélica Avendaño, ^hAlejandra García, ⁱSteffen Härtel

^aDepartment of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

^bSalud + Desarrollo Program, Corporación de Fomento de la Producción, Santiago, Chile

^cSchool of Engineering, ^dSchool of Medicine, Universidad de Concepción, Concepción, Chile

^eSchool of Medicine, Universidad de Talca, Talca, Chile

^fFaculty of Engineering, Universidad de Valparaíso, Valparaíso, Chile

^gCenter for Medical Informatics and Telemedicine, Faculty of Medicine, Universidad de Chile, Santiago, Chile

Abstract

Despite the continuous technical advancements around health information standards, a critical component to their widespread adoption involves political agreement between a diverse set of stakeholders. Countries that have addressed this issue have used diverse strategies. In this vision paper we present the path that Chile is taking to establish a national program to implement health information standards and achieve interoperability. The Chilean government established an inter-agency program to define the current interoperability situation, existing gaps, barriers, and facilitators for interoperable health information systems. As an answer to the identified issues, the government decided to fund a consortium of Chilean universities to create the National Center for Health Information Systems. This consortium should encourage the interaction between all health care stakeholders, both public and private, to advance the selection of national standards and define certification procedures for software and human resources in health information technologies.

Keywords:

Medical Informatics; Public Health Informatics; Organization and Administration

Introduction

Health care spending in Chile has grown considerably over the last years. In 2015, Chile spent 7.7% of its gross domestic product (GDP) on health care [1]. Sixty-one percent of the total spending was allocated to public health services which covers around 80% of the population. From an administrative standpoint, public health care provision is organized into 29 independent health departments, in charge of regional hospitals and secondary outpatient clinics; primary care administration is in charge of almost 300 independent municipalities [2]. A myriad of private health care providers are also available, ranging from large, multi-hospital health care networks to single physicians' private practices. There is significant cross-over movement of patients from the public to the private systems and vice versa, with some private providers having up to 50% of their patients coming from the public system [3]. From the financing perspective, working citizens are mandated to contribute at least 7% of their monthly income to health care coverage, which can be obtained through the public insurer, the National Health Fund (FONASA) or through multiple for-profit health care

insurance companies (ISAPREs). This complex organization generates a massive fragmentation of patient information.

Until 2008, the implementation of electronic health records (EHRs) was mainly driven by individual efforts at private hospitals and academic medical centers who could afford such projects [4]. However, in 2006, and after several unsuccessful attempts, the government declared the digitalization of the health care system as a key priority. The driver for that process was, for almost 10 years, a public procurement framework called Health Care Network Information System (SIDRA, for its acronym in Spanish). Through SIDRA, public health departments could buy pre-assessed health information technology solutions [5]. Although interoperability was one of the concerns when creating the SIDRA strategy, initially there was no emphasis in the adoption of standards. Moreover, the SIDRA strategy did not include private health providers, which created additional information silos.

Although to date more than 70% of all primary health clinics have adopted some kind of electronic medical record solution [6], each one of the 29 public health departments operates as an information silo incapable of communicating actionable patient information with each other. This situation becomes more complex if we take into account existent information silos among private health care providers, making it extremely difficult to monitor the population's health using health information technologies.

Within this context, strategies to advance the adoption of health information technologies have been recently given priority by the current government as a driver that could lead to economic growth and, at least in part, help reduce health care equality gaps. To support this effort, the National Development Agency (CORFO, for its acronym in Spanish) created the *Health + Development* program (S+D, for its acronym in Spanish), whose aim is to set up a public-private collaboration to establish the needs, current gaps and priorities for the health information technology sector. With the support of the RAND Corporation and different national private and public stakeholders from government, industry and universities, a roadmap emerged to streamline the development of the Chilean health information technology sector. S+D's main objective is to promote competition and innovation among health information technology providers by facilitating the adoption of standards and promoting the coordination among the different stakeholders (see Figure 1). Towards that goal, and as the first approved project, CORFO funded the creation of the National Center for Health Information Systems (CENS, for its acronym in Spanish). The

center would initially focus on interoperability, software certification, and the promotion of formal health informatics training programs to educate and certificate advanced human capital in this domain.

Public-Private coordination as a key element

The S+D program began its operation as a government agency with no prior experience in the health information technology sector. As such, finding support from the health IT experts and government health authorities was one of its first priorities. To achieve the above, the team conducted over 100 interviews and working group meetings, promoting private-public discussions. The results of these discussions were documented and validated by all participants in several published reports [7, 8, 9]. Those reports condensed the existing knowledge and perceived needs of the country's main stakeholders in this domain.

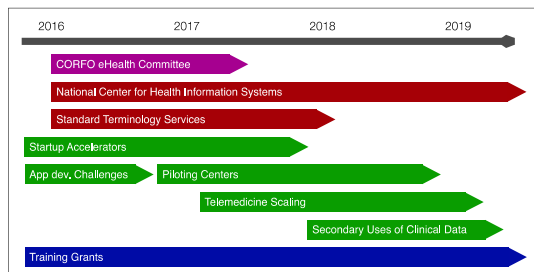


Figure 1—The creation of a National Center for Health Information Systems is part of a broader ecosystem that seeks to create a mature health information technologies market in which developers and health care providers can collaborate to create new products and services.

Additionally, and to ensure political support, the S+D program convened a diverse set of government authorities and representatives from the private sector, academia and patients to participate in its board of directors. This independent board of directors was assigned the tasks of prioritizing projects, assure the communication between the different actors of the health IT sector and, in the future, enforce the use of standards for interoperability.

After obtaining public and private commitment to the S+D strategy, the board of directors defined health care information interoperability as the main priority. In the context of a national health information technology strategy that had started almost 10 years before—which indeed advanced the adoption of electronic health records—it was now important to ensure the correct flow of patient information as a way of generating efficiencies in the use of resources and, ultimately, provide better care. Following RAND corporation's recommendations, the S+D program (together with CORFO) secured funding for the creation, through a public call for proposals, of a National Center for Health Information Systems that should largely focus on health information interoperability.

Public funding for a private non-profit corporation

One of the main issues faced by economic development agendas are the changes that occur when new governments are installed, and Chile has been no exception. A proven mechanism to provide continuity is to involve and commit the private sector or the academia in the development of such

projects. CORFO, as the agency in charge of promoting innovation and entrepreneurship has a longstanding experience in creating such alliances and was, therefore, chosen as the appropriate mechanism to fund the creation of the National Center for Health Information Systems.

Establishing such a center in the health care domain presented additional challenges. In comparison to other relevant economic activities in Chile, such as mining or agriculture, the development of the health information technology sector (according to the strategy defined by the S+D program) also had to reduce the inequality gap between private and public care. Such goal was not entirely possible to meet if standards and certifications were left in hands of the health IT industry alone, as it has been the case in previous, even successful, experiences elsewhere [10]. The decision was then to invite Chilean universities to participate in the call to create the center, and to require the establishment of a new non-for-profit organization to host the newly constituted center. It was anticipated that this decision would also help to consolidate different teams of experts currently working in the health IT domain and, therefore, strengthen national human capacities.

Finally, Universities were required to include private entities in their proposals so they could contribute to the sustainability of the center in the long run but, through established mechanisms of participation, establishing a diverse representation and ensuring objectivity and reducing conflicts of interests.

Current Advances and Proposed Organizational Structure

The request for proposals opened in July 2016 and after a review by national and international panelists the project was awarded to consortium of five Chilean universities, with support from public and private health care providers, national and international health information technology organizations, and national and international health information system vendors.

As anticipated, the center's governance is a key issue to ensure its sustained success and independence. To make sure this is the case, the center has been designed to have a board of directors constituted by members designated by the funding universities, as well as representatives from national public health entities (Chilean Ministry of Health, FONASA), software vendors and independent members (Figure 2). During the first year of the center's operations, the main task will be to establish its by-laws and constitute it as a non-for-profit entity according to the requirements established by Chilean legislation. We anticipate that this process will be completed during 2017. One of the critical components of these by-laws will be the explicit definition of the mechanisms required to add new participants to this center, with the goal to continuously ensure a broad and diverse representation.

The main lines of action proposed by CENS are the following:

- **Define interoperability standards:** the center will not recreate previously existing standards but will work with internationally recognized standard definitions and organizations to identify the best available ones for the Chilean market, and broker their adaptation to the local needs when required. To accomplish this, CENS has initiated cooperation agreements with international entities such as IHE-Europe and Salud.uv.
- **Establish software certification and testing procedures:** the center aims to establish explicit and transparent certification criteria and testing

procedures. The center will not conduct testing activities but will outsource those activities to adequate testing facilities that already operate in Chile.

- **Establish professional certification procedures:** given the limited availability of formal training programs in health informatics in Chile, the workforce is today constituted by professionals that have mostly learned on the job. Based on international experience on the subject [11], the center will establish professional profiles and certification procedures to add transparency to the job market in health informatics.
- **Technology surveillance:** the health IT domain is changing fast and will continue to do so. As a consequence, the center will establish formal technology surveillance procedures to ensure that its standards and recommendations follow internationally respected best practices.
- **Consulting:** as a vehicle to ensure transferability of the knowledge generated through its activities, the center will establish formal consulting services that will be available to all stakeholders in the Chilean health IT market.

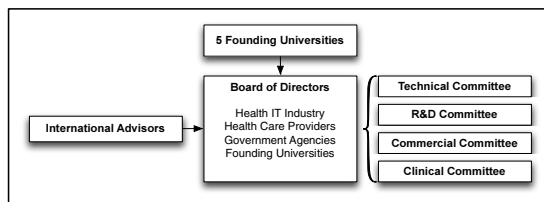


Figure 2 - The proposed organizational structure is composed by a Board of Directors with members named by founding universities as well as individuals representing a set of domain expert committees. These committees will be the setting in which most stakeholders will interact to define proposals to the Board. Additionally, the center will establish a set of external international advisors.

Conclusions

As a private-public strategy, achieving health information interoperability for the Chilean health system, remains a communication and cooperation challenge. As of today the success of the predefined projects and roadmap lies mainly within the “social capital” that was built by Salud+Desarrollo. Although there is broad consensus among stakeholders on the path to follow and the steps needed to do so successfully, a critical aspect is the sustained political commitment of S+D’s board of directors. Chile is facing a presidential election during 2017, and CENS will have to consolidate its role in the ecosystem and become the public space for keeping that social capital. To do so, the social capital and public trust needs to be soon transformed into concrete deliverables that prove to be useful for all actors in the Chilean health care system.

At the same time, the independent nature of the entity requires a business model for sustainability that has to consider both local and international markets, making the need to participate in the international fora, highly relevant.

Finally, the recent changes in the digitalization strategy of the Ministry of Health, changes that include the end of SIDRA as a mechanism for buying health information system solutions, creates new demand for internal capabilities since public

health departments will have to create their own frameworks at the same time they are required to abide by new interoperability standards.

Acknowledgements

This work has been funded by the Chilean Development Agency (CORFO) grant number 16CTTS-66390.

References

- [1] Health Statistics 2016 [Internet] Organization for Economic Cooperation and Development. 2016 [cited 21 December 2016]. Available from <http://www.oecd.org/els/health-systems/health-data.htm>.
- [2] Missoni E, Solimano G. Towards Universal health coverage: the Chilean experience. World Health Report (2010), Background Paper 4. [Internet] World Health Organization; 2010 [cited December 21, 2016]. Available from: <http://www.who.int/healthsystems/topics/financing/healthreport/4Chile.pdf?ua=1>
- [3] Dimensionamiento del Sector de Salud Privado en Chile. [Internet] Clínicas de Chile. 2014 [cited December 21, 2016]. Available from http://www.clinicasdechile.cl/wpcontent/uploads/2016/04/Dimensionamiento_sectorprivado_cifras2014.pdf
- [4] Capurro D. Health informatics in Chile: responding to health reforms. Health Information & Libraries Journal. 2007 Dec 1;24(4):287-91.
- [5] SIDRA – Sistemas Información de Red Asistencial [Internet] Ministerio de Salud, Chile; 2015 [cited December 21, 2016]. Available from <http://www.minsal.cl/SIDRA>
- [6] Estrategia Digital en Salud: Avance y Proyección SIDRA [Internet]. Unidad de Modernización y Gobierno Digital. 2015 [cited 22 December 2016]. Available from: http://www.modernizacion.gob.cl/site/assets/files/1225/avance_estrategia_digital_salud_2015_gabriela_lissi.pdf
- [7] A Roadmap for the Development of Health Information Technology in Chile [Internet]. Santa Monica, CA: RAND Corporation, 2016. Available from http://www.rand.org/pubs/research_reports/RR1358z2.html.
- [8] Tecnología y Servicios [Internet]. Santiago, Chile: Salud + Desarrollo; 2016. [cited December 21, 2016]. Available from: <https://static1.squarespace.com/static/55dc9136e4b05820bf029511/u/5779bbf5f7e0ab3baff559b3/1467595770100/Informe+Descripción+de+la+industria.pdf>
- [9] Informe Hoja de Ruta Programa Estratégico Nacional Salud + Desarrollo [Internet]. Santiago, Chile: Salud + Desarrollo; 2016. [cited December 21, 2016]. Available from: <https://static1.squarespace.com/static/55dc9136e4b05820bf029511/u/5779bbf5f7e0ab3baff559b3/1467595770100/Informe+mesas+de+trabajo+2016.pdf>
- [10] Health Information Technology - What was CCHIT? [Internet] 2016. Certification Commission for Health Information Technologies [cited December 21, 2016]. Available from <https://www.cchit.org>
- [11] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, Hovenga E, Lun KC, Marin H, Martin-Sanchez F, Wright G. IMIA Recommendations on Education Task Force Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. First revision. Methods Inf Med. 2010 Jan 7;49(2):105-20.

Address for correspondence

Daniel Capurro, MD, PhD

Diagonal Paraguay 362, 5° Piso, Santiago, Chile

dcapurro@uc.cl

A Combined Collaborative Information Behaviour (CIB) and Continuity of Care Framework for Modeling Complexity in Colorectal Cancer Screening Access

Caitlin Champion^{a,b}, Ewan Affleck^c, Gonzalo G. Alvarez^{a,d}, Craig Kuziemsky^b

^aFaculty of Medicine, University of Ottawa, Canada

^bTelfer School of Management, University of Ottawa, Canada

^cNorthwest Territories Health and Social Services

^dSchool of Epidemiology, Public Health and Preventative Medicine, University of Ottawa, Canada

Abstract

Colorectal cancer screening access within a rural and remote health care environment represents a complex systems problem. Existing modeling approaches are inadequate in their representation of health system complexity. A combined Collaborative Information Behavior (CIB) and Continuity of Care framework was developed to model the health care processes involved in screening access over time. This framework highlighted necessary information behavior supports and system gaps in screening access, supporting development of targeted informatics solutions to improve screening access and cancer outcomes.

Keywords:

Public Health Informatics; Models, Theoretical; Continuity of Patient Care

Introduction

Health care systems are complex adaptive systems, and understanding health system interrelationships and context may provide insights that are not readily apparent when studying system components in isolation [1,2]. In conceptualizing health system complexity, modeling can provide important insights into health system interrelationships and impacts of proposed health transformation initiatives aimed at improving quality of care [3,4]. Qualitative modeling is a necessary initial step, as complex processes must be clearly understood before they can be effectively translated into more formal modeling approaches [5,6,7].

Existing health care process modeling approaches adapted from Business Process Management literature (e.g. Workflow patterns or languages such as Unified Modeling Language or Business Process Modeling Notation) are limited in their ability to capture health system complexity, as they are often specific to data or people, assume that processes are static, and are unable to describe interrelationships between them [5,7,8,19]. Knowledge intensive processes and connected communities bring with them added complexity that cannot be modeled by rigid control flows or by assuming a common frame of reference about how participants will behave in certain roles [19, 20].

Information generation, utilization and management form much of the basis of health care. Understanding relationships between health information systems and its users and processes is necessary for successful implementation [3,9]. However, information management processes in healthcare are often dynamic and connected over time and across different tasks and providers [14].

The Northwest Territories is a rural and remote region of Canada with a large indigenous population where colorectal cancer incidence is high, outcomes are poor, and screening rates are lower than other parts of the country [10, 11]. Access to colorectal cancer screening has been identified as a strategic priority in the territory for improving screening rates and patient outcomes [12].

Colorectal cancer screening access in the Northwest Territories represents a complex health system problem, as it involves cyclical multidisciplinary health care processes within a distributed health care environment occurring over many years. Health system complexity may contribute to health outcome inequities among marginalized rural and remote, and indigenous populations [13]. The ability to effectively model health system complexity is a necessary first step to develop and implement health transformation initiatives to improve colorectal cancer screening access and patient outcomes within complex health care settings like the Northwest Territories.

As described above, existing modeling approaches often do not account for dynamic complex processes that take place over time. New modeling approaches that capture the interrelationships between people and information over time is necessary to understand health system complexity and develop targeted health transformation initiatives, such as Electronic Medical Records (EMRs) that can improve quality of care.

The objective of this paper is to address the above need by developing a qualitative modeling approach integrating Collaborative Information Behaviour (CIB) and the Continuity of Care frameworks. The integrated framework allows us to describe and understand the complex interrelationships between people and information within colorectal cancer screening access and how they evolve over time.

Methods

Conceptual Framework

Collaborative Information Behaviour (CIB) provides a framework for describing information behavior within a multidisciplinary health care team, categorizing health care processes in the form of health system agents, interactions and information problems and their resulting system behaviors influenced by system context [14]. It is triggered by high problem complexity, lack of immediate information access, lack of domain expertise, and fragmentation of information resources, all of which are present in colorectal cancer screening in the Northwest Territories [14].

Continuity of Care, as a cornerstone of care access, provides a framework for conceptualizing health care access over time,

and is defined as “the degree to which a series of discrete healthcare events is experienced as coherent and connected, and consistent with the patient’s medical needs and personal context” [15,16]. It captures information flow (informational continuity), management across multiple health care providers (management continuity), and patients and provider relationships (relational continuity) within care access [16].

Where CIB describes information behavior within a multidisciplinary health team at a single point in time, Continuity of Care describes health system interrelationships contributing to long-term care access. Combined, these models were used to provide a framework for conceptualizing interrelationships between the people and information involved in colorectal cancer screening access over time.

Data Collection and Analysis

Semi-structured interviews with health care providers involved in colorectal cancer screening across the Northwest Territories, including primary care providers, endoscopists and endoscopy scheduling providers (N=29) were performed using CIB and continuity of care categories to guide discussion. A subset of interview transcripts (N=22) from three health authorities was selected for in-depth qualitative content analysis as a representative sample of territorial screening processes and used to construct qualitative health system models of colorectal cancer screening access within a combined CIB/continuity of care framework [17].

Data analysis involved coding of verbatim interview transcripts and subsequent content analysis of coded passages following coding methods of Saldana [18] and qualitative content analysis methods of Miles, Huberman and Saldana [17]. A logbook was used to record data analysis decisions and the progression of analysis in the generation of study results. First round coding was conducted concurrently with the data collection phase, while further code refinement and data analysis occurred following the completion of data collection. A codebook was iteratively developed through regular analysis meetings between research team members throughout the analysis process with analysis decisions being recorded in the logbook.

Participant validation of health system models was performed through review of exploratory process models by one to two study participants from the study. Community presentations of general study results outlining colorectal cancer screening access problems and contextual factors influencing screening access were also completed as an alternate method of participant validation engaging both study participants and health system stakeholders.

Results

Qualitative Health System Model Construction

Information behavior, or the utilization of health care information within colorectal cancer screening, ultimately links all health system components. Information behavior is generated by interactions between health system components, including health system agents and decision nodes, and forms the health system interrelationships across the continuity of care streams of relational continuity (patient-provider interactions), management continuity (provider-provider interactions), and informational continuity (provider-information system

interactions) that make up colorectal cancer screening access. In our study, information behavior is used to represent health system activities. By examining information behavior within the exploratory process models, it is apparent that health system information processes enable information flow through the health system to support the patient interactions required for colorectal cancer screening access.

Qualitative health system models of colorectal cancer screening access were developed as swimlane diagrams. Fig. 1 provides an example of one model and shows an example of information behavior generated from a primary care encounter interaction (relational continuity) between patients and providers (health system agents), subsequently generating the decision node sequence of screening initiation, risk stratification, and FIT initiation. Fig. 1 also highlights the complexity in colorectal cancer screening. Records of previous care may be obtained from historic documents (management continuity) or retrieved from the health centre chart (informational continuity) that inform the decision node of screening initiation for the current primary care encounter. The decision node sequence generated by the encounter is then documented in the primary care encounter note (informs management continuity) and subsequently stored in the health centre chart (informational continuity). As demonstrated by this sequence of health care processes, information flow from patient interactions through decision nodes, management continuity, and informational continuity is necessary to enable screening access.

Swimlanes were used to show sequential activities and responsibilities over time and to organize continuity of care categories of relational continuity, managerial continuity and informational continuity, with CIB components of health system agents (people and information systems) within each swimlane [8,14,16]. The model was developed according to health system processes described by participants in the interviews.

The CIB problems, labelled as decision nodes within the model, formed a link between relational continuity, shown as interactions between patients and providers, and managerial continuity, shown as provider-provider interactions. Patient-provider interactions identified health issues needing to be addressed in screening, shown as decision nodes, which generated information behaviors by providers to address the need, either by another provider or the same provider later in the colorectal cancer screening sequence. These links between different providers or the same provider at different times within the health system were achieved through medical record documents such as referrals, encounter notes, or consultation notes. Information behaviors were identified as links between health system agents, decisions and documents within the health system.

Health System Complexity in Colorectal Cancer Screening Access

The combined CIB/Continuity of Care framework provided two types of insight. First was the identification of information behavior issues to provide insight on screening access. Second was identification of contextual factors that impact screening access.

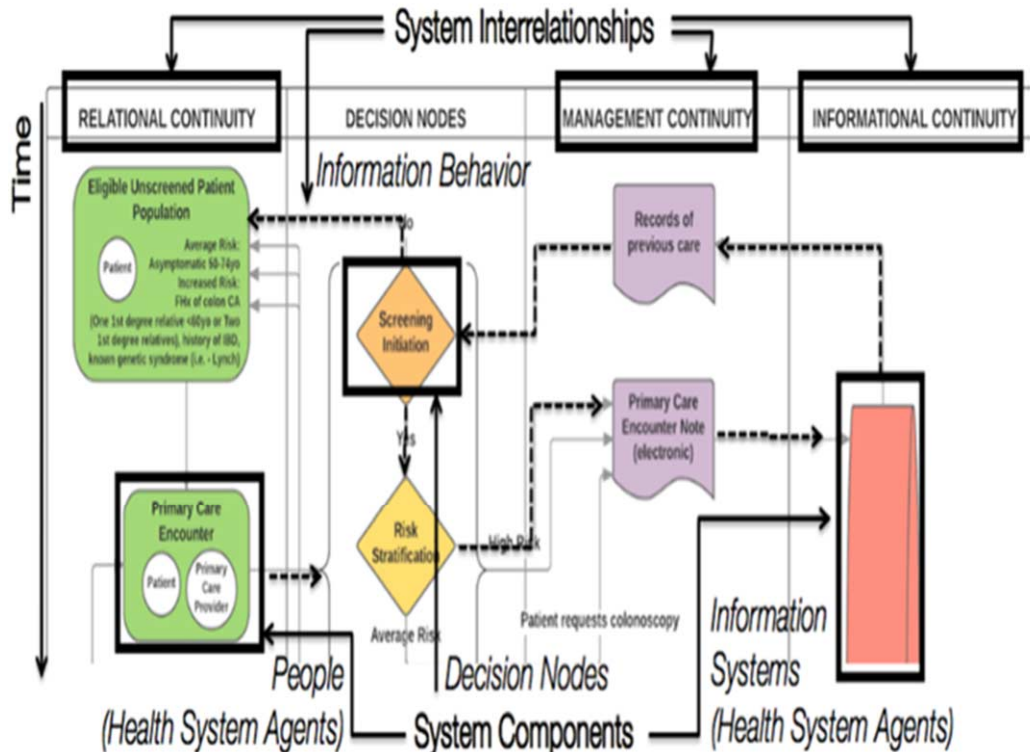


Figure 1 - Qualitative Health System Model Using CIB/Continuity of Care Framework

Information Behavior Issues Impacting Screening Access

Information behaviors were identified as foundational health system interrelationships involved in colorectal cancer screening access can identify areas where information behaviors may be optimized to improve screening access. Information behavior gaps in colorectal cancer screening access identified within models included poor quality referral documentation, scheduling complexity, inconsistent results reconciliation and follow-up documentation, siloed information systems and discontinuity in screening recall.

An example of information behavior issues in scheduling complexity is whether appointments are coupled or uncoupled. Coupled pre-colonoscopy consultation and colonoscopy appointments are scheduled concurrently to occur over a short timeframe (i.e. – single travel encounter) compared to uncoupled appointments, which are scheduled separately over a longer timeframe and may require multiple travel encounters. Another information behavior issue leading to system complexity within colonoscopy screening access was the number of patient and provider interactions at the time of pre-colonoscopy consultation and colonoscopy procedure and the travel required on behalf of patients and endoscopists. System complexity in colonoscopy screening access increased with the number of providers involved, as well as with increased travel requirements among both patients and providers.

Finally, patient information throughout the territory is often siloed, increasing the risk of missed or duplicate information, and requiring providers to access multiple information sources to access the information required to support decision making. As a result of siloed information some providers described getting triple copies of some results and having to keep manual records to track what reports they have seen.

Contextual Factors Impacting Screening Access

The combined CIB/continuity of care framework also identified contextual factors impacting screening access including geographic factors such as the need for medical travel, social factors such as patient awareness and residential school experiences, and health system factors such as rebooking forgiveness and locum-based practice influencing colorectal cancer screening access. For example, a locum practice environment may make screening initiation more difficult due to lack of familiarity with territorial practice guidelines, as well as lack of continuity with individual patients. Poor attention to social determinants of health including job insecurity, housing insecurity, and food insecurity, may also present a barrier to screening initiation, as patients may be less interested in preventative care when their basic needs are unmet.

Discussion

While systems modeling is an important tool in health system transformation, established Business Process Management modelling approaches are limited in their ability to fully characterize health system complexity such as dynamic and connected processes [5, 19, 20]. This paper addressed that issue and presented a combined CIB/continuity of care framework to identify and describe complex health system interrelationships and information behaviour gaps. Our modeling framework provides two key type of insight. First it enables us to understand patient flows through the health system over time through relational continuity, but also the information behaviour interrelationships required to facilitate the management and informational continuity aspects of care

access. Second, our framework also helps identify contextual issues that may further impair information behaviour interrelationships. We provided proof of concept of our framework using a case study of colorectal cancer screening access in the Northwest Territories.

Conceptualizing the complexity of information behaviour interrelationships in colorectal cancer screening access helps identify areas of excess complexity where streamlining of processes is required to prevent system failure, or areas of low complexity where additional system supports are required. To support colorectal cancer screening access in rural and remote health system environments such as the Northwest Territories, targeted informatics tools such as endoscopy referral forms, reporting forms to support recall reconciliation, and enterprise information systems may support improved screening access by supporting sequential information behaviors within the health system. However, these informatics solutions must be considered in light of contextual factors such as geographic, social and health system factors.

Another benefit of modeling frameworks is that they provide a visual representation of complex health system processes. The systems models generated through our combined CIB and Continuity of Care modeling framework can promote common understanding of the complexity of colorectal cancer screening access among health system stakeholders including physicians, other care providers and policy-makers.

Next stages of our research are translation of these qualitative health system models into a system design framework for quantitative modeling to support quantitative measurement of the potential impacts of proposed health system transformation initiatives.

Limitations of this research is that our study took place in one context and other CIB and Continuity of Care issues may emerge in other settings. Our modeling approach may be applied to other health system access problems to inform targeted health informatics solutions to optimize care access.

Conclusions

A combined CIB and continuity of care framework can be applied to qualitative modeling of colorectal cancer screening access to conceptualize health system complexity. This framework supports understanding of health system interrelationships including information behaviour complexity and associated gaps to identify targeted solutions such as health informatics tools for improving care access.

Acknowledgements

We thank the University of Ottawa Division of General Surgery Surgeon Scientist Program and University of Ottawa Clinician Investigator Program for their financial support and encouragement. Project funding was made possible through a CIHR Frederick Banting and Charles Best Canadian Graduate Scholarship and Student Research Grant from the Telfer School of Management Research Fund

References

- [1] T.G. Kannampallil, G.F. Schauer, T. Cohen, V.L. Patel, Considering complexity in healthcare systems. *Journal of Biomedical Informatics* **44** (2011), 943-947.

- [2] J.P. Sturmberg, D.M. O'Halloran, C.M. Martin, Understanding health system reform - a complex adaptive systems perspective. *Journal of Evaluation in Clinical Practice* **18** (2012), 202-208.
- [3] E. Coiera, *Guide to Health Informatics Third Edition*. Boca Raton: CRC Press, 2015.
- [4] D.L. Cooke, Y. Yang, G. Curry, P.R. Rogers, T.R. Rohleder, R.C. Lee, D. Strong, Introducing System Dynamics Modelling to Health Care in Alberta. In *Proceedings of the 25th International Conference of the System Dynamics Society*, Boston, July 2007.
- [5] A.A. Rad, M. Benyoucef, C.E. Kuziemsky, An Evaluation Framework for Business Process Modelling Languages in Healthcare. *Journal of Theoretical and Applied Electronic Commerce Research* **4**(2) (2009), 1-19.
- [6] M. Roberts, L.B. Russell, D. Paltiel, M. Chambers, P. McEwan, M. Krahm, Conceptualizing a Model: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force - 2. *Medical Decision Making* **32** (2012), 678-689.
- [7] C. Kuziemsky, Decision-making in healthcare as a complex adaptive system. *Healthcare Management Forum* **29**(1) (2016), 4-7.
- [8] G.T. Jun, J. Ward, Z. Morris, J. Clarkson, Health care process modelling: which method when? *International Journal for Quality in Health Care* **21**(3) (2009), 214-224.
- [9] J.S. Ash, M. Berg, E. Coiera, Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors. *JAMIA* **11**(2) (2004), 104-112.
- [10] Northwest Territories Department of Health and Social Services. Cancer in the Northwest Territories 2001-2010. Yellowknife: NWT Department of Health and Social Services, (2014).
- [11] H. Singh, C.N. Bernstein, J.N. Samadder, R. Ahmed, Screening rates for colorectal cancer in Canada: a cross-sectional study. *CMAJ Open* **3**(2) (2015), E149-E157.
- [12] Northwest Territories Department of Health and Social Services. (2015) *Charting Our Course: Northwest Territories Cancer Strategy 2015-2025*. Yellowknife: NWT Department of Health and Social Services.
- [13] S. Hill, D. Sarfati, B. Robson, T. Blakely, Indigenous inequalities in cancer: what role for health care? *ANZ Journal of Surgery* **83** (2013), 36-41.
- [14] M.C. Reddy, B.J. Jansen, P.R. Spence, P.R. Collaborative Information Behaviour: Exploring Collaboration and Coordination during Information Seeking and Retrieval Activities. In J. Foster (Ed.) *Collaborative Information Behaviour: User engagement and communication sharing*. Hershey, PA: IGI Global, 2010.
- [15] Canadian Medical Association. Health Care Transformation in Canada. Ottawa: Canadian Medical Association. (2010)
- [16] J. Haggarty, R.J. Reid, G.K. Freeman, B.H. Starfield, C.E. Adair, R. McKendry, Continuity of care: a multidisciplinary review. *British Medical Journal* **327** (2003), 1219-1221.
- [17] M.B. Miles, A.M. Huberman, J. Saldana, *Qualitative Data Analysis 3rd edition*. Los Angeles: Sage Publications, 2014.
- [18] J. Saldana, *The Coding Manual for Qualitative Researchers*. Los Angeles: Sage Publications, 2012.
- [19] G. Bruno, F. Dengler, B. Jennings et al., Key challenges for enabling agile BPM with social software. *Journal of Software Maintenance and Evolution: Research and Practice* **23** (2011), 297-326.
- [20] N. Russell, A. Barros, Business Processes in Connected Communities. In: Fournier F., Mendling J. (eds) *Business Process Management Workshops. BPM 2014. Lecture Notes in Business Information Processing*, vol **202**. (2015) Springer, Cham, 446-451.

Address for correspondence

Corresponding author: Dr. Craig Kuziemsky, Telfer School of Management, University of Ottawa, 55 Laurier Avenue East, Ottawa, ON K1N 6N5; Email: kuziemsky@telfer.uottawa.ca

Clinical Decision Support and Primary Care Acceptance of Genomic Medicine

Dian A. Chase^a, Sherry Baron^b, Joan S. Ash^a

^aOregon Health & Science University, Portland, OR, USA,

^bCity University of New York, New York, NY, USA

Abstract

Clinical decision support systems (CDS) have an important role in the implementation of precision medicine, particularly for pharmacogenomics. This study examines potential factors for their acceptance by primary care clinicians.

For this qualitative study we purposively selected five U.S. primary care sites with a variety of sizes, electronic health record vendors, and patients. We interviewed an average of seven clinicians per site.

Clinicians placed a low priority on incorporating pharmacogenomics into practice. Other themes included the potential of precision medicine, clinician unfamiliarity with genomics, minimal evidence for primary care uses, additional costs and time burdens, workload, and a need to first successfully complete other electronic health record interventions.

This study outlines issues in implementing primary care precision medicine and the role for genomic CDS. Currently there are significant barriers. With more evidence and the development of effective CDS, however, there is potential for turning each of the barriers into facilitators.

Keywords:

Decision Support Systems; Clinical, Precision Medicine; Genomics

Introduction

Genomics-guided precision medicine – the ability to individualize therapies and guidance based on an individual’s genetic makeup – has been a research target for over fifty years [1]. While there have been significant breakthroughs in oncology, primary care applications of genomics precision medicine have been limited [2]. But the potential is there. Just two of the CYP450 enzymes are involved in the metabolism of ninety percent of all drugs [3]. The expression of these enzymes can be predicted using genomic data, and patients divided into fast, normal, slow and non-metabolizers for a given drug. How an individual metabolizes a drug can determine whether a “safe” dose represents an ineffective treatment, an effective treatment, or an overdose.

Clinical trials are underway to demonstrate the value of using genomics guided precision medicine in primary care settings – ClinicalTrials.gov lists several hundred trials with potential applications in primary care [4]. Once demonstrated however, the deployment of this genomics based precision medicine in primary care will still face challenges. Effective CDS can increase the usage of genomic testing for primary care. Peterson et. al. [5], found that two-thirds of clinicians who had a genomics CDS available planned to order a genomics-based

test, compared to only 13% in an earlier survey of those without a robust CDS [6].

This qualitative study, based on interviews with clinicians at five very different sites, examines some of the barriers to clinical decision support systems for precision medicine.

Methods

This study was conducted within the context of a larger study on the use of clinical decision support for workers’ health issues [7]. Five U.S. sites were selected based on maximum differences in geography, experience with using an EHR, and type of parent health care organization. All of the sites were implementing the patient-centered medical home delivery model. Subjects were selected with the assistance of an inside contact person based on their roles as clinicians (MDs, DOs, PAs, and NPs).

The data were collected using a Rapid Assessment Process (RAP) [8-9]. A multi-disciplinary team visited each of the sites and conducted both semi-structured interviews and observed clinicians using the EHR. The process utilized a set of structured tools (interview guides, observation templates, and overall clinic data); results and progress were reviewed by the team every evening, and presented to the clinic at the end of our visits.

The data presented here are largely based on semi-structured interviews, as we did not observe any use of genomics data in the clinics we visited. To start the conversation clinicians were asked “How would you prioritize the implementation of meaningful use, work related data, and genomic medicine CDS?” For those who asked for clarification we provided the example of pharmacogenetic dosing for coumadin. After they replied, we would ask them to elaborate. Probes included questions about their knowledge of genomic based precision medicine and what barriers they perceived to its implementation. Their responses were recorded and transcribed.

We followed guidelines for assuring rigor by triangulating (multidisciplinary researchers, multiple sites, different types of data), member checking (feedback to sites about results), auditing (tracking data gathering), reflexivity (researchers’ awareness of their own bias), and data saturation (gathering data until little more is being learned) [10]. Data were analyzed with the assistance of qualitative data analysis software in three ways. First, a template method was used to identify answers to our interview questions. Second, a grounded hermeneutic approach was taken to discover patterns and themes across sites. Finally, comments about the factors influencing precision medicine implementation and their connection to CDS were extracted, coded, grouped by themes, and interpreted.

Approval was obtained from the OHSU and NIOSH Institutional Review Boards as well as the IRB's at the sites we visited.

Results

Participating clinics

Five sites participated. An average of seven primary care clinicians were interviewed at each site for this portion of the study.

The number of providers (Doctors of Medicine and Osteopathy, Nurse Practitioners, and Physician Assistants) at the sites selected varied from 5 to over 2,000; the geographic locations spanned both East and West Coasts and from the Gulf of Mexico to the Great Lakes; the type varied from community clinics to large academic medical centers, and the experience with EHRs varied from three to fifteen years. The EHR vendors included NextGen, AllScripts, and EPIC.

Themes

1. Priority for genomic CDS

Although the potential of genomic medicine was a motivating factor for the inclusion of genomic-related CDS for the most of the clinicians, none of those interviewed viewed genomic CDS as a higher priority than including other data, including work-related CDS, in the EHR. Factors that influenced these lower priorities were the interviewee's personal unfamiliarity with genomic medicine, a lack of compelling evidence for better outcomes, the increased cost of testing, potential work flow interruptions, and the need for other, higher priority, EHR related interventions. Clinicians across all sites were intrigued by the concept, but did not believe genomic-based precision medicine was ready for wide application.

2. Perceived potential

The concept of individualized medicine was appealing to the interviewees. As one said: *"I'm a little wary of it but I'm also excited because I think it will help us get closer to having an individualized plan for people that really is about who they are."* This knowledge could then translate into more effective treatments: *"if there's a way to pinpoint what medication would really be the best for the patient based ... genetic makeup or, you know, whatever is going on, then I think that would be great... I have seen a lot of patients where we try one, two or three different types of IV treatments and it just doesn't work and you have to wait so many months for it to kick in and if it doesn't, then switch over... if there's a way to narrow it down to ... find ... that best drug."* And they also thought it might help with risk assessment *"I think it's important to try to identify future risks for people, not just-- which you can get at with cancer risks but-- so what are the future risks of you developing diabetes?"* This positive attitude was consistent across clinics.

3. Clinician lack of knowledge

There was a wide range of familiarity with genomic medicine, from those who have worked in the field to those who profess ignorance: *"we did some work when we were using Naltrexone for ... so I'd be interested."* *"Well, genomic information. Okay. So there's -- and maybe I don't know enough about it. ... I don't think there's a specific place that I would find it except I refer a lot of people to medical geneticists."* *"I mean I manage a lot of patients, you know, each month and I don't know anything about, you know, whether their genes tell me to do a certain thing."* Although

our samples were not selected to allow quantitative analysis, it did appear that the larger clinics were more likely to have practitioners who were at least aware of current work in genomics.

4. Need for evidence

Some thought that they were getting enough "genetic data" by just reviewing the family history, while others cited a lack of trials showing significant benefits for primary care providers: *"We do that already. Like, if the patient when they come in for physicals, we ask them family history."* *"nobody has actually linked any of those genetic markers to anything that we do."* This pattern appeared to be consistent across clinics.

5. Cost concerns

Not only were the additional tests viewed as adding additional cost for limited value, there was also the questions of who would pay for them: *"It's the cost containment. I probably would not [use genomic testing] if there wasn't a huge difference in patient outcomes."* *"I don't even know if MediCal [state insurance] covers things [DNA testing for breast cancers] like this."* Cost concerns appeared to be more frequently measured by the community clinics.

6. Work flow issues

Without a significant perceived benefit, providers were unwilling to have interruptions on their workflow: *"So that would involve additional blood work ... I'm not sure how that would be incorporated into my work flow."* *"It is not part of the flow."* This was a concern for all clinics, and appeared to be more prevalent for clinicians with prior EHR implementation experience.

7. Current EHR implementation and priorities

Even if the providers wanted to incorporate genomic information into their CDS, there are many priorities that were viewed as more important, including work-related data: *"We can't do everything...right now the priority is to make the system work for us I think we are working too hard for it."* *"I think [incorporating occupational health and social determinants of health] can help more right now than personalized medicine and at lower cost."* But there is a possibility that this could change in the future: *"I probably don't think about genomics, ... but in the future, if we have that kind of information, that would be awesome."* Once again, this was a uniform concern across all clinics.

Discussion

This is the first multi-site qualitative study of this topic that spans community, regional, and academic health care sites. In addition to the barriers identified in prior studies on genomic-based CDS, we found that the status of the current EHR system also affected the prognosis for genomics based CDS. Our results also differed from more generalized CDS implementation studies – the lack of evidence and provider knowledge created barriers that are more significant than for other types of CDS.

As discussed above, we identified six factors involved in the acceptance of genomic-based CDS: the potential for improving care, clinician knowledge about precision medicine, evidence that genomic based CDS improves outcomes, costs, workflow integration, and other priorities due to system status.

These factors are similar to those found in two earlier studies (by Unertl et al [11] and Haga et al [12]), both of which were conducted at one site and had a lower number of primary care

clinicians participating. The study by Unertl was conducted as semi-structured interviews at a large academic medical center following the implementation of CDS for genomic based dosing of clopidogrel and coumadin. The Haga findings were based on three focus groups at another academic medical center that had not implemented a genomic CDS. The Unertl study included six primary care physicians and nine cardiologists, the Haga study twelve primary care providers and six geneticists. Table 1 compares results of the three studies.

Table 1 – Comparison of factors for success

Theme	This study	Unertl et al [11]	Haga et al [12]
Potential of Genomic Based Precision Medicine	x	x	x
Clinician Knowledge	x	x	x
Need for Strong Evidence	x	x	
Costs	x	x	
Workflow	x	x	x
Current EHR Implementation	x		

Some of these barriers have been identified in prior work on general CDS implementation. Both Bates' Ten commandments for successful CDS [13] and Kawamoto's review of factors for successful CDS [14] implementation cite workflow as an important element in CDS success. Only seven percent of all CDS implementations, however, had the evidence included as one of its components [14], despite the strong association with implementation success. Further, only 31% of all CDS efforts had an associated educational component; the inclusion of which was not associated with increased success [14]. We believe that genomics implementations are more likely to require an educational element based on this studies providers' knowledge issues.

A limitation of our work is that it was conducted in parallel with a study of the inclusion of worker-based health CDS in the EHR and this could have biased our results. But, as shown in Table 1, our results are similar to those of two prior studies. The other two studies also had two themes ours did not – the need for help in communicating results to patients and issues with data storage and usage after the initial test. This may be due to a difference in the degree of implementation of genomics testing. Our study included sites that had not implemented a robust primary care genomic CDS, the other two studies either involved medical geneticists or already had a robust implementation.

Another limitation of our work is that it was a qualitative study; our work was designed to discover the range of possible responses. As a result, any of the differences between clinics is not statistically valid and only a starting point for future work.

Of the six factors shown in Table 1, one is a facilitator – clinicians can see a potential benefit. The other five are currently barriers. But these barriers can be changed to facilitators. Clinical trials can, and hopefully will, produce clear evidence of better outcomes. Once these trials are

completed effective CDS can help with the other barriers, especially those involving clinician readiness. Our results lead to the following recommendations:

Clinician knowledge: Effective clinical decision support can provide “just in time” knowledge to clinicians through best practice alerts before the patient encounter, alerts at the time orders are placed, pop-ups with additional information regarding labs, and linkages to more information. Continuing education of clinicians about genomic medicine is needed for increasing their awareness of the evidence as well.

Costs: With effective data transfer from other institutions a genomic test can be done once and shared across multiple organizations, thus minimizing costs. There are already CDS implementations that have significantly reduced duplicate lab test, similar strategies can be used for genomic testing [15].

Workflow: Gathering the genomic data can be implemented by adding the typical tests required to the normal order sets for new patients and incorporating them into the templates. It is encouraging to note that workflow considerations are part of several current pharmacogenomics CDS projects [5,16].

Communication: Customized patient handouts can assist clinicians by providing simple illustrations of probabilistic concepts. These handouts can help implement shared decision making using graphical tools (see <http://shareddecisions.mayoclinic.org>).

Current EHR Implementation: And as the EHR implementations mature and organizations become more proficient at implementing CDS, the development and deployment of genomic CDS will become easier.

Conclusion

Except for the potential of genomic-based precision medicine, the factors identified in our interviews are currently barriers to its acceptance. With more evidence, and education, CDS will be a critical tool in bringing the promise of genomic guided precision medicine to fruition. The barriers – clinician knowledge, need for evidence, costs, workflow, and other EHR issues can be addressed with CDS.

The primary care clinicians we interviewed, from a wide range of clinics, want to improve the care they provide for patients. They only need to see a clear path to do so. What is needed is an easy to use, relatively low cost CDS intervention with clear evidence of improved patient outcomes for primary care. There are many clinical trials currently underway that are attempting to provide the evidence to support the CDS, but successful implementation will require both the evidence and the tools to implement. Clinical decision support is needed to make these interventions less expensive, easier for both clinicians and patients to understand, and able to fit into the clinical workflows.

Acknowledgements

This work was partially supported by CDC/NIOSH Contract 200-2015-61837 as part of NORA project #927ZLDN. We would like to gratefully acknowledge the contributions of Margaret S. Filios, M.Sc, R.N., Stacey Marovich, M.S., M.H.I., Jane Wiesen, Ph.D., and Genevieve B. Luensman, Ph.D. to this work.

References

- [1] M. Ingelman-Sundberg, Pharmacogenetics of cytochrome P450 and its applications in drug therapy: the past, present, and future, *Trends Pharm Sci*, **25** (2004), 193-200.
- [2] J. Moaddeb and S. Haga, Pharmacogenetic testing: current evidence of clinical utility, *Ther Adv Drug Saf*, **4** (2013), 155-169.
- [3] I. Vizirianakis, Advances of Pharmacogenomics toward phrmacotyping in drug prescription, in *Handbook of Personalized Medicine: Advances in Nanotechnology, Drug Delivery, and Therapy*, I. Vizirianakis ed, Pan Stanford Publishing, Singapore, 2013.
- [4] U.S. National Institutes of Health, www.clinicaltrials.gov, accessed November 27, 2016
- [5] J.F. Peterson, J.R. Field, J.S. Schildcrout, J.C. Denny, T.L. McGregory, S.L. VanDriest, et al, Attitudes of clinicians following large-scale pharmacogenomics implementation, *Pharmacogenomics J* **16** (2016), 393-398
- [6] E.J. Stanek, C.L. Sanders, K.A. Johansen Taber, M. Khalid, A.Patel, R.R Verbrugge, et al, *Clin Pharm Ther* **91** (2012), 450-458
- [7] J.S. Ash, D. Chase, J.F. Wiesen, E.V. Murphy, and S. Marovich, Studying Readiness for Clinical Decision Support for Worker Health Using the Rapid Assessment Process and Mixed Methods Interviews, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2016, p. 285.
- [8] J. Beebe, *Rapid Assessment Process: An Introduction*. Walnut Creek, CA, AltaMira Press, 2001.
- [9] C.K. McMullen, J.S. Ash, D.F. Sittig, A. Bunce, K. Guappone, R. Dykstra, J. Carpenter, J. Richardson, A. Wright. Rapid assessment of clinical information systems in the healthcare setting: An efficient method for time-pressed evaluation. *Meth Inform Med* **50** (2011), 299-307.
- [10] B.F. Crabtree, W.L. Miller, *Doing qualitative research*, 2nd edition, Sage Publications, Thousand Oaks, CA, 1999
- [11] K.M. Unertl, J.R. Field, L. Price, J.F. Peterson, Clinical perspectives on using pharmacogenomics in clinical practice, *Per Med* **12** (2015), 339-347
- [12] S. Haga, G. Tindall, and J.M. O'Daniel, Professional perspectives about pharmacogenetic testing and managing ancillary findings, *Genet Test Mol Bio Ma* **16** (2012), 21-24
- [13] D.W. Bates, G.J. Kuperman, S. Wang, et al. Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality. *Journal of the American Medical Informatics Association : J Amer Med Inform Assoc* **10** (2003), 523-530
- [14] K. Kawamoto, C.A. Houlihan, E.A. Balas, and D.F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success, *BMJ* **330** (2005), 765.
- [15] G.W. Procop, L.M. Yerian, A.M Harrison, K. Kottke-Marchant, Duplicate laboratory test reduction using a clinical decision support tool, *Am J Clin Pathol* **141** (2014), 718-23
- [16] A. Burcur, J. vanLeeuwen, N. Christodoulou, K. Sigdel, L. Koumakis, N. Graf, G. Stamatakis, Workflow-driven clinical decision support for personalized oncology, *BMC Med Inf Dec Mak* **16** (2016), 87

Address for correspondence

Dian Chase
chased@ohsu.edu
 Oregon Health & Science University
 3455 S.W. Veterans Hospital Road
 Portland, Oregon, USA, 97230

An Experimental Comparison of Quality Models for Health Data De-Identification

Johanna Eicher, Klaus A. Kuhn, Fabian Prasser

Institute of Medical Statistics and Epidemiology, University Hospital rechts der Isar, Technical University of Munich, Germany

Abstract

When individual-level health data are shared in biomedical research, the privacy of patients must be protected. This is typically achieved by data de-identification methods, which transform data in such a way that formal privacy requirements are met. In the process, it is important to minimize the loss of information to maintain data quality. Although several models have been proposed for measuring this aspect, it remains unclear which model is best suited for which application. We have therefore performed an extensive experimental comparison. We first implemented several common quality models into the ARX de-identification tool for biomedical data. We then used each model to de-identify a patient discharge dataset covering almost 4 million cases and outputs were analyzed to measure the impact of different quality models on real-world applications. Our results show that different models are best suited for specific applications, but that one model (Non-Uniform Entropy) is particularly well suited for general-purpose use.

Keywords:

Privacy, Personally identifiable information, Data anonymization

Introduction

The collaborative collection and sharing of sensitive individual-level data has become an important aspect of modern biomedical research. The secondary use of health data for research purposes is a typical example [1]. To protect privacy in such scenarios, a broad spectrum of safeguards must be implemented, including data use agreements and fine-grained access control [2]. On the data-level, anonymization is a central safeguard. There are various ways and definitions. One important aspect is data *de-identification*, which focuses on protecting data from re-identification. For this purpose, datasets are transformed in such a way that it becomes very difficult to link their records to identified individuals without investing a disproportionate amount of time and effort [3].

There are *rule-based* and *computational* approaches to the de-identification of health data. The Safe Harbor method of the US Health Insurance Portability and Accountability Act (HIPAA) [4] is a typical example for the former type. It specifies 18 transformation rules which describe the removal or alteration of attribute values that are associated with a high risk of re-identification (e.g. names and dates). In other jurisdictions, e.g. in Germany [3], regulations are less interpretable and computational methods to data de-identification are thus more important. Here, data is transformed (semi-) automatically to ensure that privacy risks are minimized.

The transformation of data inevitably leads to loss of information. Therefore, a balance has to be sought between an increase in privacy protection on one side and a decrease in data

quality on the other. *Privacy models* and *quality models* are used to quantify the two aspects. The contradiction between the two conflicting optimization goals is typically resolved by specifying a risk threshold for the privacy model. This reduces the de-identification process to a simpler optimization problem, in which the objective is to make sure that risk thresholds are met while data quality is maximized [2].

Objective

Measuring data quality is a non-trivial issue as the nature of usefulness of data often depends on the use case [5]. As it is typically unknown in advance how the data will be analyzed, models are needed, which quantify data quality for general-purpose use. Fung et al. proposed to measure the *similarity* (which can be defined in multiple ways) between the original and the de-identified data [6]. Domingo-Ferrer et al. noted that a quality model should capture the amount of information loss for a reasonable range of data uses [7]. They introduced two characteristics, *analytically valid* and *analytically interesting*, which need to be present for a dataset to have *little* loss of information. In this context, analytical validity requires the preservation of certain statistical characteristics, while data is said to be analytically interesting if some useful attributes for further analyses remain intact [7].

A wide variety of *general-purpose quality models*, which aim to distinguish valid or interesting data from invalid or uninteresting data, have been proposed and used in scientific papers. Typically, these models define a decrease in data quality, as well as an increase in information loss, which can be quantified [5]. The notion of using information loss as an indicator for data quality is also prevalent in official statistics, namely the so-called *score*, which measures the trade-off between quality (information loss) and privacy (disclosure risk of the released data) [8]. Even though various papers have compared data de-identification algorithms, a systematic evaluation of quality models has not been conducted yet. Consequently, a guideline for selecting appropriate models for specific scenarios is missing. Potential application scenarios for de-identified data include the privacy-preserving sharing of data from research registries or health databases. De-identified data extracts may also be used to provide partners with an overview of data, which is potentially available for sharing in a fine-grained form. Finally, de-identified data can also directly be used for advanced analytics and observational research, e.g. for building predictive models.

As a first step towards the development of a guideline, we have implemented and evaluated several general-purpose quality models with the intention of answering the following questions:

1. How do common models for measuring data quality influence the way in which datasets are transformed?
2. If different models are used, how are the obtained results related to each other?

- How well is de-identified data, obtained by using different quality models, suited for real-world applications?

Methods

Background

In data de-identification, the general attack vector assumed is *linkage* of a sensitive dataset with an identified dataset (or similar background knowledge about individuals). *Identity disclosure* (or *re-identification*) means that an individual is successfully linked to a specific data record [9]. This is a very important type of privacy breach, as it has legal consequences for data owners according to many laws and regulations worldwide. As a first step towards data de-identification, *directly identifying* information (such as names) must be removed [10]. The remaining attributes, which may be used for linkage, are termed *quasi-identifiers* (or indirect identifiers, or keys). Such attributes are not directly identifying, but they may be used in combination for linkage. It is further assumed that they cannot simply be removed from a dataset, as they may be required for analyses and that corresponding information is likely to be available to an attacker.



Figure 1 – Generalization hierarchies for “age” and “sex”

When data is de-identified, values of quasi-identifiers are transformed to ensure that the data fulfills privacy requirements. This can be performed with user-defined *generalization hierarchies* [11]. Examples are shown in Figure 1. Here, values of the attribute “age” are transformed into intervals, with decreasing precision on increasing levels of generalization. Values of the attribute “sex” can only be suppressed. Generalization hierarchies are well suited for categorical attributes, but they can also be constructed for continuous attributes through categorization.

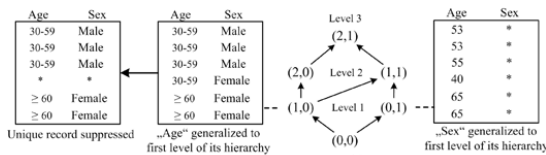


Figure 2 – Example showing different transformations represented as a generalization lattice

In order to transform the data, *globally-optimal full-domain anonymization* algorithms have been recommended. Such algorithms construct a search space in a structure called *generalization lattice*. An example for the lattice constructed from the hierarchies from Figure 1 is shown in Figure 2. The graph displays each node that represents a single transformation, which defines generalization levels for all quasi-identifiers. An arrow denotes that a transformation is a direct *generalization* of a more *specialized* transformation, which means that it increments exactly one generalization level as defined by its predecessor. The original dataset (0, 0) is at the bottom, whereas the transformation with maximal generalization (2, 1) is at the top. The search space is then

traversed to find a transformation, which results in output data that fulfills all privacy requirements, and at the same time provides optimal data quality.

Protection against re-identification is often implemented with the *k-anonymity* privacy model [9]. A dataset is *k-anonymous* if, regarding the quasi-identifiers, each record cannot be distinguished from at least $k - 1$ other records. This property can be used to define *equivalence classes* of indistinguishable records [12]. The basic idea is that an attacker will only be able to associate any individual with at least k records, which reduces the probability of correct linkage to not more than $\frac{1}{k}$. The left-most output dataset in Figure 2 fulfills 2-anonymity, which means the maximal re-identification risk of any record is 50%.

In order to create de-identified datasets of high quality, attribute generalization can be combined with the suppression of data records. This means that records from equivalence classes, which violate the privacy model (i.e. *outliers*), are automatically replaced with semantic-free placeholders. Because of record suppression, less generalization is required to ensure that the remaining records fulfill the privacy model, which increases the quality of de-identified datasets [13]. In the left-most dataset from Figure 2, one record has been suppressed in the output of applying the transformation (1, 0).

The aforementioned quality models are used to rank the different privacy-preserving output datasets and to select a transformation that maximizes data quality. The *Loss* model, for example, measures data granularity by analyzing the extent to which the domain of an attribute is covered by the transformed values [14]. For the datasets shown in Figure 2, the reduction in data granularity is 30%, 16% and 50%, from left to right.

Quality Models

We have implemented the following quality models (QMs) into ARX, which is an open-source de-identification tool for biomedical data [13]:

- *Average Equivalence Class Size (AECS)* is a row-oriented model, which measures the average size of equivalence classes of indistinguishable records [15].
- *Discernability* is a row-oriented model, which measures the size of equivalence classes combined with a penalty for suppressed records [16].
- *Precision* is a cell-oriented model, which quantifies data quality by reporting the amount of distortion of attribute values. Distortion is measured as the generalization level of an attribute value relative to the height of the attribute's generalization hierarchy [17].
- *Loss* is a cell-oriented model, which measures the granularity of data by determining the fraction of an attribute's domain that is covered by the transformed values [14].
- *Ambiguity* is a row-oriented model, which measures the degree of uncertainty in the resulting data [18].
- *Kullback-Leibler (K.-L.) Divergence* is a row-oriented model, which measures differences in the distributions of equivalence class sizes [19].
- *Non-Uniform (N.-U.) Entropy* is a column-oriented model, which measures differences in the distributions of attribute values induced by data transformations [20]. It is based on the concept of mutual information, which quantifies the amount of information that can be obtained about one variable by observing the other.

Dataset

Table 1 – Description of the patient discharge dataset

Attribute	Type	Description
Hospital ID	Spatial	A unique identifier
Age	Demographic	Patient's age at admission in years
Sex	Demographic	Patient's sex
Ethnicity	Demographic	Patient's ethnicity
Race	Demographic	Patient's racial background
ZIP Code	Spatial	Patient's ZIP code of residence
County	Spatial	Patient's county of residence
Length of stay	Temporal	Total number of days from admission to discharge
Admission quarter	Temporal	The calendar quarter the patient was admitted
Charge	Sensitive	Total charges for the stay

We used each quality model to de-identify a publicly available patient discharge dataset [21]. The dataset contains 3,985,166 records and 10 attributes (Table 1). In our experiments, we have defined all spatial, demographic and temporal attributes as quasi-identifiers. As we will describe later, we used the remaining sensitive attribute (charge) for determining the usefulness of output data.

Privacy Protection

We de-identified the dataset with attribute generalization and record suppression to produce output datasets, which fulfill the *k-anonymity* privacy model. We chose $k=5$, which is a typical parameter in the biomedical domain that specifies a re-identification risk of not more than 20% for each record [2].

Experimental Design

We addressed the first research question, i.e. how quality models influence the way in which datasets are transformed, analyzing how much generalization and record suppression had to be used in the de-identification process to achieve optimal data quality. The former is expressed as a *generalization degree* for each attribute, which is defined as the relative generalization level to which it was transformed. The latter is expressed as the number of *removed records*.

To answer the second question, i.e. how the results obtained with different models are related to each other, we used each model to assess the quality of the optimal solutions obtained with all other models. To make the different quantifications of quality comparable to each other, we normalized them: a value of 0% represents the original data and a value of 100% represents a dataset where all information has been removed.

Finally, to answer the third question, i.e. how well the de-identified data is suited for real-world applications, we analyzed the impact of the different methods of data transformation on typical use cases. Moreover, we employed

statistical classification, which is a common application scenario for individual-level data [22]. The aim was to predict the values of a selected *class attribute* from a set of *feature attributes*. This is implemented with supervised learning, where a model is created from a training set. We used the discharge dataset to build logistic regression models [23], which were able to predict the height of the bill for hospital stays, i.e. whether the *charge* for a stay was below \$10,000, between \$10,000 and \$50,000, or greater than \$50,000.

To be able to quantify the analytical validity of the de-identified data, we created classifiers, which could be evaluated using the original input data; although they have only been trained with de-identified output data. For this purpose, we implemented the approach presented in [22] into ARX. For evaluating different predictors, we used 10-fold cross-validation. We normalized all resulting prediction accuracies into the range [0, 1], where 0% represents the accuracy of the trivial *ZeroR* method, which simply always returns the most frequent class value from the original dataset [23], and 100% represents the accuracy of a logistic regression model trained with the original, unmodified input dataset.

Results

How do common models for measuring data quality influence the way in which datasets are transformed?

Table 2 – Generalization degrees and removed records (RR).

QM	Generalization degrees	RR
AECS	5·100%, 4·0%	25%
Disc.	6·100%, 1·57%, 2·0%	0%
Precision	1·100%, 1·60%, 1·33%, 6·0%	7%
Loss	1·67%, 1·60%, 1·57%, 1·33%, 5·0%	5%
Ambiguity	4·100%, 1·67%, 1·57%, 1·50%, 1·17%, 1·0%	0%
K.-L. Div.	2·100%, 1·17%, 6·0%	21%
N.-U. Ent.	4·100%, 1·71%, 1·50%, 3·0%	10%

Table 2 shows the generalization degrees, and the number of removed records for the outputs obtained by de-identifying the discharge dataset with each quality model.

It can be seen that the fraction of removed records varied between 0% and 25%. With each quality model, at least one attribute was preserved as-is, while just the result obtained with the Loss model did not contain at least one completely generalized attribute. The dataset was transformed with high degrees of generalization when AECS, Discernibility, Ambiguity or N.-U were used. Entropy was used to quantify the loss of information. Just little generalization was used when quality was measured with Precision, Loss and K.-L. Divergence. No records were removed when Discernibility or Ambiguity were used, while a large proportion of the records was removed when quality was measured with AECS and K.-L. Divergence. With the models Precision, Loss and N.-U. Entropy, the dataset was transformed with a balanced combination of both attribute generalization and record suppression.

Table 3 – Relative information loss in percent

QM used for evaluation	QM used for de-identification						
	AECS	Discernibility	Precision	Loss	Ambiguity	K.-L. Divergence	N.-U. Entropy
AECS	0.0004	0.0096	0.0019	0.0024	0.0930	0.0011	0.0011
Discernibility	24.5553	0.0390	6.9322	5.4202	0.0424	20.5476	9.6655

Precision	66.4688	72.8116	26.9124	28.0155	62.9714	39.6713	62.0003
Loss	66.4688	68.4609	18.4723	9.3869	49.3907	38.2668	56.4513
Ambiguity	24.5547	0.0280	6.9172	5.3887	0.0070	20.5426	9.6612
K.-L. Divergence	83.8570	96.7961	45.6947	49.2434	72.6321	27.8739	55.1117
N.-U. Entropy	62.4520	65.9717	48.9954	51.8416	81.0103	44.4787	39.4218

How are the datasets obtained with different models related to each other?

Table 3 shows how the different models assessed the quality of output data obtained using the other models. Each column represents the result of de-identifying the data with a single model as indicated. In each row, a model was used to assess the quality of the output obtained with the other models. Consequently, the highlighted values on the diagonal represent the optimum for each model.

It can be seen that, in terms of AECS, all results had comparable data quality. However, information loss was considered very low in general. When using Discernibility and Ambiguity, results obtained with the other models were determined to be much worse and quality values differed by orders of magnitude. When using Precision and Loss, quality values were within a reasonable range, considering the transformations, which were applied to the data (Table 2). However, both models, as well as K.-L. Divergence, measured big differences between the different solutions. When using N.-U. Entropy, the quality of the results from the different models was placed in a reasonable range, and different solutions were considered to be of rather comparable quality.

How well is de-identified data obtained with different quality models suited for real-world applications?

Before building the prediction models, we performed a feature selection process. The results showed that neither the age of a patient nor the length of a stay was predictive for the prices charged by the hospitals. Therefore, we built classifiers, which predicted the charge from the spatial features *hospital-ID* and *county of residence*, the demographic parameters *sex*, *ethnicity* and *race*, as well as temporal information in form of the *admission quarter*.

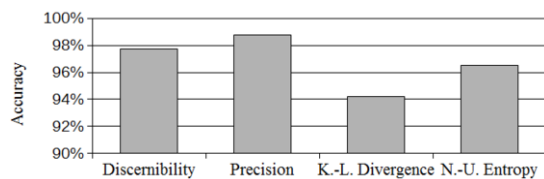


Figure 3—Relative accuracies of logistic regression models

The results obtained by training models with the output of using AECS, Loss and Ambiguity performed not very well with prediction accuracies below 30%. Figure 3 shows the results obtained using the remaining quality models. It can be seen that when using Discernibility, Precision, K.-L. Divergence and N.-U. Entropy, the de-identification process had just negligible effects on the performance of the prediction models; we measured relative accuracies between 94% and 98%. This means that the models performed almost as well as models trained with unmodified input data.

Discussion

Our experiments indicate that different models are suited best for different application scenarios. When using the AECS model, datasets were de-identified with a high degree of generalization and a high degree of record suppression. Moreover, predictive models created from the output obtained

with AECS exhibited sub-optimal performance. This shows that the model is not suited well for real-world applications in biomedicine. When using the models Discernibility or Ambiguity, datasets were de-identified with attribute generalization only. This means that the models are suited well for de-identifying small datasets, e.g. from rare disease networks or data collections from sparsely populated regions, where statistical power may otherwise be reduced disproportionately. Using the models Precision, Loss or N.-U. Entropy resulted in a balanced application of both attribute generalization and record suppression. This means that the output data is well suited for providing potential data sharing partners with an overview of available data, as instance-level and schema-level information is preserved. Finally, the models Discernibility, Precision, K.-L. Divergence and N.-U. Entropy are suited well for de-identifying data that is to be used for predictive modeling. The latter two models are based on stringent information theoretic foundations, and it is thus not surprising that output obtained with them is suited well for machine learning purposes. In contrast, we did not expect to obtain such good results when using Discernibility and Precision, as both are rather simple in nature.

Our results have also shown that the utility or usefulness of data does not necessarily correlate with the degree of quality measured by general-purpose models. In future work, we also plan to investigate *special-purpose quality models*, which are models that have been designed with specific usage scenarios in mind.

The application scenario investigated in this article, statistical classification, is a well-known example of a specific application scenario. While our results have shown that data de-identified with general-purpose quality models can be suited well for this context, specialized quality models also have been proposed. They minimize the loss of information for features, which are most discriminating for a specified class attribute [6]. This has been shown to optimize output data for classification purposes [16].

Another application scenario is the de-identification of diagnosis codes for use in association studies between phenotypic and genotypic data [24]. The transactional characteristics of such data require that irrelevant inter-attribute relationships are removed, which can be achieved with specialized de-identification algorithms that also require specific data quality models. In future work, we plan to investigate such models, e.g. utility constraints [24] as well.

Conclusion

Non-Uniform Entropy is a quality model, which has frequently been recommended for de-identifying health data, e.g. by Emam et al. [25]. Based on the results of our experiments, we can confirm that the model provides the best results for general-purpose usage. With this model, de-identified data contained instance-level and schema-level information. Moreover, statistical power was reduced by only 10%. Finally, by using de-identified data with optimal quality according to this model as a training set, we were able to build a statistical classifier with good prediction accuracy.

References

- [1] J. Christoph, L. Griebel and I. Leb, Secure secondary use of clinical data with cloud-based NLP services, *Methods of Information in Medicine* **54** (2015), 276–282.
- [2] K. El Emam and B. Malin, Appendix B: Concepts and methods for de-identifying clinical trial data, *Sharing clinical trial data: maximizing benefits, minimizing risk*, The National Academies Press, 2015.
- [3] Federal Data Protection Act, version promulgated on 14 January 2003 (Federal Law Gazette I p. 66), as most recently amended by Article 1 of the Act of 14 August 2009 (Federal Law Gazette I p. 2814).
- [4] U.S. health insurance portability and accountability act of 1996, Public Law, 1996, p. 1349
- [5] B.C.M. Fung, K. Wang, A.W.-C. Fu and P.S. Yu, *Introduction to privacy-preserving data publishing: Concepts and techniques*, CRC Press, 1st ed., 2010, ISBN 9781420091489.
- [6] B.C.M. Fung, K. Wang, R. Chen and P.S. Yu, Privacy-preserving data publishing: A survey of recent developments, *ACM Computing Surveys* **42** (2010), 14.
- [7] J. Domingo-Ferrer, Sánchez and S. Hajian, Database privacy, S. Zeadally and M. Badra, editors, *Privacy in a digital, networked world: Technologies, implications and solutions*, Springer, 2015.
- [8] J. Nin, J. Herranz and V. Torra, Towards a more realistic disclosure risk assessment, *Privacy in Statistical Databases*, Springer, 2008, vol. 5262 of Lecture Notes in Computer Science, 152–165.
- [9] L. Sweeney, *Computational disclosure control - A primer on data privacy protection*, Ph.D. thesis, Massachusetts Institute of Technology, 2001.
- [10] F. Prasser, R. Bild, J. Eicher, H. Spengler, F. Kohlmayer and K.A. Kuhn, Lightning: Utility-driven anonymization of high-dimensional data, *Transactions on Data Privacy* **9** (2016), 161–185.
- [11] P. Samarati, Protecting respondents' identities in microdata release, *Transactions on Knowledge and Data Engineering* **13** (2001), 1010–1027.
- [12] P. Samarati and L. Sweeney, Generalizing data to provide anonymity when disclosing information, in *Symposium on Principles of Database Systems (ACM)* (1998).
- [13] F. Prasser, F. Kohlmayer, R. Lautenschläger and K.A. Kuhn, ARX – A comprehensive tool for anonymizing biomedical data, *AMIA Annual Symposium Proceedings, AMIA*, 2014, 984–993
- [14] V.S. Iyengar, Transforming data to satisfy privacy constraints, *International Conference on Knowledge Discovery and Data Mining*, ACM, 2002, 279–288.
- [15] K. LeFevre, D.J. DeWitt and R. Ramakrishnan, Mondrian multidimensional k-anonymity, *International Conference on Data Engineering*, IEEE, 2006, 25.
- [16] R.J. Bayardo and R. Agrawal, Data privacy through optimal k-anonymization, *International Conference on Data Engineering*, IEEE, 2005, 217–228.
- [17] L. Sweeney, Achieving k-anonymity privacy protection using generalization and suppression, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10** (2002), 571–588.
- [18] M.E. Nergiz and C. Clifton, Thoughts on k-anonymization, *International Conference on Data Engineering*, IEEE, 2006, 96.
- [19] A. Machanavajjhala, D. Kifer, J. Gehrke and M. Venkatasubramanian, l-diversity: Privacy beyond k-anonymity, *Transactions on Knowledge Discovery from Data* **1** (2007), 24–35.
- [20] A. De Waal and L. Willenborg, Information loss through global recoding and local suppression, *Netherlands Official Statistics* **14** (1999), 17–20.
- [21] D. Sanchez, S. Martinez and J. Domingo-Ferrer, Comment on "Unique in the shopping mall: On the reidentifiability of credit card metadata" (Supplementary Materials: <http://arxiv.org/abs/1511.05957>, Archive: http://crises-deim.urv.cat/opendata/SPD_Science.zip), *Science* **6279** (2016), 1274–1274.
- [22] A. Inan, M. Kantarcioglu and E. Bertino, Using anonymized data for classification, *Proceedings - International Conference on Data Engineering* (2009), 429–440.
- [23] I.H. Witten and F. Eibe, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005.
- [24] G. Loukides, A. Gkoulalas-Divanis and B. Malin, Anonymization of electronic medical records for validating genome-wide association studies., *Proceedings of the National Academy of Sciences of the United States of America* **107** (2010), 7898–7903.
- [25] K. El Emam and L. Arbuckle, *Anonymizing health data: Case studies and methods to get you started*, O'Reilly Media, Inc., 1st ed., 2013, ISBN 978-1-449-36307-9.

Address for correspondence

Fabian Prasser, Chair of Medical Informatics, Institute of Medical Statistics and Epidemiology, University Hospital rechts der Isar, Technical University of Munich, Ismaninger Str. 22, 81675 Munich, Germany; E-mail: fabian.prasser@tum.de.

Evaluation of the Association Between Different Patient Indexing Strategies and Effective Indexing in Health Centers of the Public System of the Autonomous City of Buenos Aires: An Exploratory Study

Santiago Esteban^a, Cecilia Palermo^a, Leandro Alassia^a, M. Victoria Giussi^a,
Analía Baum^a, Fernán González Bernaldo de Quirós^b

^a Clinical informatics, Statistics and Epidemiology Office, Health Ministry, Ciudad Autónoma de Buenos Aires, Argentina

^b Government Health Advisor, Ciudad Autónoma de Buenos Aires, Argentina

Abstract

During the implementation of an electronic health record (EHR) system in the public system of the city of Buenos Aires, Argentina, different patient indexing strategies were devised and implemented. We sought to assess the association between these strategies and effective indexing (proportion of patients who are indexed and have a consultation registered in the EHR). Strategies were grouped into three modalities (High, Intermediate, and Low intensity). We estimated hazard ratios (HR) comparing the High and Low intensity to the Intermediate strategies. The crude analyses showed a significant difference between the curves ($p < 0.0001$). In the multivariate analysis, the HRs of High and Low intensity interventions showed on average, values above 1 from 0 to 90 days compared to the Intermediate intensity strategy (High: 2.08 (1.65, 2.52); Low: 2.59 (2.29, 2.9)). From that point on, the HRs of both strategies were not different from 1.

Keywords:

Electronic Health Records

Introduction

During the implementation of an electronic health record (EHR) system, the processes of identity accreditation and the creation of a master patient index are critical in order to avoid duplication or misassignment of a person's clinical data [1-3]. In turn, depending on the implementation context, these two processes may present different levels of complexity, which conditions the strategies used to promote patient indexing [4; 5].

Beginning in June 2016, the Ministry of Health of the Autonomous City of Buenos Aires, through the Office of Clinical Informatics, Statistics and Epidemiology (OCISE), started implementing the computerization of all health records from the public health system [6]. This involves the development and implementation of an EHR, initially covering outpatient clinics but also more complex levels of care, such as hospitals. In this context, the patient indexing process is particularly difficult. Factors related to physical resources, human resources, and characteristics of the population of each center condition the indexing methodology, making the implementation of a single, common strategy for all centers, impossible.

The reason why different strategies coexist is because each primary care health center is virtually autonomous. Our team

had to negotiate with the chairmen of each center on how to implement the patient indexing process. Thus, a particular indexing strategy was devised by the chair of each primary care health center taking into account the characteristics of the center and its population. With the aim of improving the future implementations, we decided to evaluate the association between the different indexing implementation strategies and effective patient indexing.

Methods

General Design and Data Source

This study utilized a prospective cohort design based on secondary data extracted from the EHR system of the city of Buenos Aires. We used the term 'prospective' because the intervention was defined and recorded prior to the occurrence of the result.

Population

All persons enrolled in the computerized health system (SIGEHOS) of the city of Buenos Aires, between 1/6/2016 and 11/24/2016, were included. Two health centers were excluded: one was already implemented prior to the start of the current process (prior to June 2016); the other, the implementation process has just begun (November 2016).

Intervention

During the first weeks of implementation in each center, sociologists made ethnographic observations [7] of the patient indexing process dynamics and held extensive and recurrent interviews [8] with the administrative staff involved in the process. From the information obtained, typologies [9] were constructed with a qualitative approach [10] in order to classify, structure, order, and compare the different conceptualizations of the indexing strategies [11; 12]. The differences in the strategies and their intensities gave rise to 3 types:

- *Low Intensity:* The registration is taken as an alternative instance to the usual process of attention since the use of the paper medical record predominates. The indexing process depends on the time availability of the administrative staff. Paper and EHRs coexist.

- *Intermediate Intensity*: Indexing is offered to patients who request new appointments and to those who visit the center requesting maternal formula or need to fill in paperwork related to social security. Paper and EHRs coexist.
- *High Intensity*: Registration is proposed as a condition in all instances of consultation at the center and the EHR is the main registration system.

Outcome

We used the time from registration to the first visit recorded in the EHR or administrative censorship (24/11/2016 or maximum of five months). This was done to account for the fact that the raw indexing total (total number of indexed patients) can reflect many patients who are indexed without actually needing to see someone from the health staff (e.g., in several centers, the handout of maternal formula was used as an instance to promote indexing). This situation is a problem for patients who need medical attention since they may decide to skip indexing because of the long waiting times and queues, thus promoting paper the perpetuation of the paper records. Given the short duration of the study and the type of population (outpatient population), no competing risks such as death were considered.

Covariates

Baseline variables were extracted from the EHR at the time of each patient's indexing, such as sex, age, type of housing, district of residence, programmatic area in which they were registered, number of professionals of the main specialties at the center of attention (Obstetrics, tocoginecology, pediatrics, family medicine), and number administrative staff.

Model Structure

Figure 1 shows the directed acyclic graph (DAG) of the defined structure to try to resolve confounding between exposure (A = indexing strategy) and the result (Y = first visit recorded in the EHR) model. In these graphs, time runs from left to right, lines denote association (bidirectional), arrows indicate causal direction and the boxes around the variables reflect controlling by that variable and therefore the rupture of the association flow through that path.

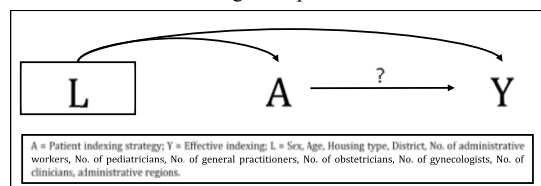


Figure 1 - DAG for the fully adjusted model

Statistical Analysis

The crude proportions of effective indexing were estimated using the Kaplan-Meier method and compared by means of the log-rank test. The level of significance for all tests was set at 0.05. A Cox proportional hazards model was used for multivariate adjustment. The proportionality of risk was assessed by analyzing the Schoenfeld residuals (graphic and test analysis) and also by means of the log-log graphs for each variable. The variables for which the assumptions did not hold were incorporated into the model through an interaction with a flexible function of time (natural cubic spline). Non-

significant interaction terms were dropped from the model. All analyzes were performed using R (R Foundation for Statistical Computing, Vienna, Austria URL.: <https://www.R-project.org>).

Results

Figure 2 shows the distribution of health centers implemented in the city of Buenos Aires. The color indicates the indexing strategy used.

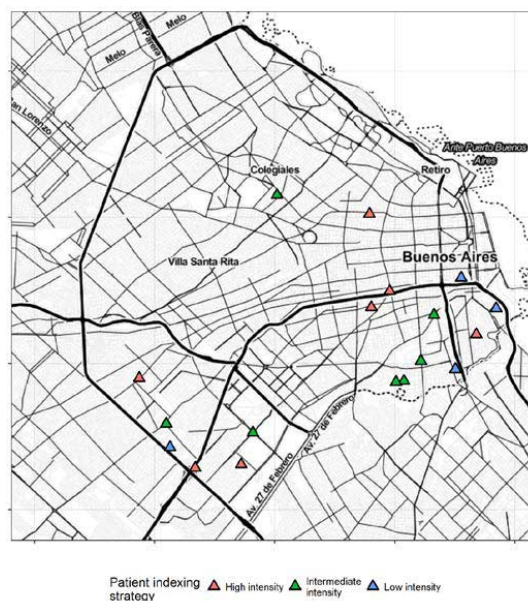


Figure 2- Map of the city of Buenos Aires with the implemented primary care health centers by indexing strategy

Table 1 shows the baseline characteristics of the patients. The distribution of age categories represents a characteristic broad-based pyramid with a clear predominance of women. In turn, the vast majority of patients reported residing in Buenos Aires (87%). 49% of the indexed patients came from four of the 18 centers analyzed (Hosp. Grierson, 5, 7 and 35). This is due to a combination of implementation time, size of the center in terms of population served, and indexing strategy. On the other hand, since the process began in the southwest area of the city, two programmatic areas of the six, account for 72% of the registered patients.

Survival Curves

The cumulative probability curves for the three strategies are plotted in Figure 3. The three strategies initially show a rapid ascent, due to those who register and have a consultation on the same day or the following day. From that moment on, the curves diverge; the High intensity is associated with a shorter time until the first visit registered in the EHR. Intermediate and Low intensity strategies initially differ, but each reaches similar levels after 60 days. The log-rank test showed statistically significant differences between the curves ($p < 0.0001$).

Table 1- Patients' baseline characteristics

	High intensity	Intermediate intensity	Low intensity	p
No. of patients	14547	22974	8581	
Sex: Female	9590 (65.9)	14740 (64.2)	5651 (65.9)	<0.001
Age (median [IQR])	16.76 [5.33, 34.15]	15.40 [5.08, 32.49]	14.32 [4.72, 31.41]	<0.001
Age categories in years (%)				<0.001
0 - 10	5633 (38.7)	9110 (39.7)	3612 (42.1)	
11 - 20	2252 (15.5)	4051 (17.6)	1316 (15.3)	
21 - 30	2241 (15.4)	3332 (14.5)	1323 (15.4)	
31 - 40	1705 (11.7)	2631 (11.5)	1028 (12.0)	
41 - 50	1129 (7.8)	1707 (7.4)	600 (7.0)	
51 - 60	745 (5.1)	1102 (4.8)	356 (4.1)	
61 - 70	478 (3.3)	667 (2.9)	224 (2.6)	
71 - 80	267 (1.8)	280 (1.2)	84 (1.0)	
81 - 90	92 (0.6)	88 (0.4)	37 (0.4)	
>90	5 (0.0)	6 (0.0)	1 (0.0)	
Residence: City of Buenos Aires (%)	11873 (82.3)	21254 (94.4)	6958 (81.5)	<0.001
Informal housing (%)	2763 (19.0)	12026 (52.3)	1419 (16.5)	<0.001
Administrative regions (%)				<0.001
Argerich	384 (2.6)	0 (0.0)	2315 (27.0)	
Durand	0 (0.0)	647 (2.8)	0 (0.0)	
Grierson	0 (0.0)	7202 (31.3)	0 (0.0)	
Penna	2484 (17.1)	9446 (41.1)	972 (11.3)	
Ramos	2151 (14.8)	0 (0.0)	0 (0.0)	
Santojanni	9528 (65.5)	5679 (24.7)	5294 (61.7)	
Effective indexing (%)	8401 (57.8)	11796 (51.3)	4432 (51.6)	<0.001

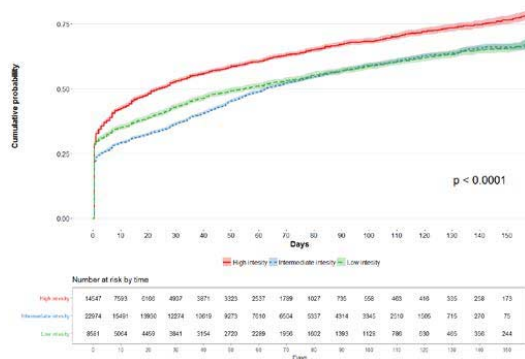


Figure 3 - Curve of cumulative probability

Proportional Hazards Assessment

All variables were analyzed using the procedures described in the methods section. The scaled Schoenfeld residuals test was highly significant for many of the variables included in the model; however, graphical analysis in many cases did not show significant deviations from proportionality. This was

probably due to the large sample size for which minimal differences were detected. The intervention variable also showed no proportionality of hazards (figure 4).

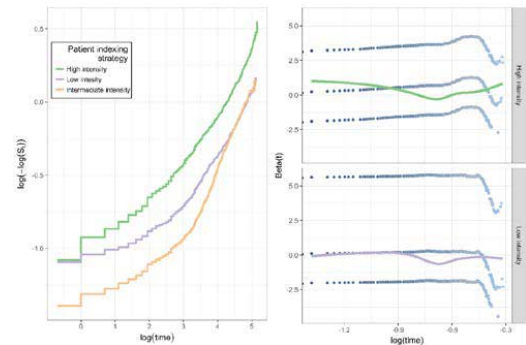


Figure 4 - Assessment of the proportional hazards assumption. Log-log and scaled Schoenfeld residuals graphs

To resolve this situation, the follow-up time was divided into seven-day intervals and the risk of each intervention (High and Low) vs the reference (Intermediate) were calculated. Within each interval the risks proved to be proportional. Figure 5A shows the progression of the hazard ratios of the high and Low intensity strategies vs Intermediate intensity, for the crude model. An initial benefit is observed approximately during the first 30 days in favour of the High and Low intensity strategies. The adjusted model included sex, district of residence, type of housing, number of administrative staff at the indexing health center, number of professionals at the indexing health center (medical clinic, pediatrics, obstetrics, tocoginecologia, familiar medicine), and programmatic area of the indexing health center. It was observed that the hazard ratios of the Low and High intensity strategies lose significance after approximately 90 days from the starting point when compared against the Intermediate intensity strategy (Figure 5B, Table 2).

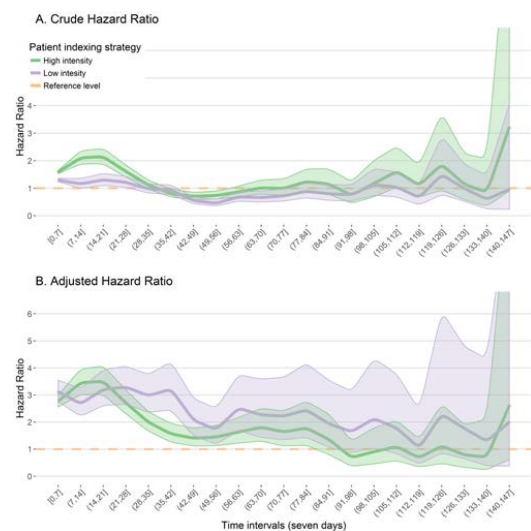


Figure 5 A&B - Hazard ratios of High and Low intensity strategies compared to Intermediate intensity

Table 2 - Hazard ratios for each time interval per intervention strategy

Time intervals (days)	High intensity HR (95% CI)	Low intensity HR (95% CI)
[0,7]	2.77 (2.56 ,3)	3.13 (2.76 ,3.55)
(7,14]	3.46 (3.02 ,3.96)	2.68 (2.21 ,3.24)
(14,21]	3.51 (3.01 ,4.08)	3.21 (2.62 ,3.93)
(21,28]	2.72 (2.31 ,3.19)	3.3 (2.67 ,4.07)
(28,35]	1.99 (1.67 ,2.38)	2.98 (2.36 ,3.76)
(35,42]	1.57 (1.26 ,1.95)	3.22 (2.45 ,4.22)
(42,49]	1.41 (1.11 ,1.78)	2.07 (1.49 ,2.87)
(49,56]	1.45 (1.13 ,1.87)	1.73 (1.19 ,2.51)
(56,63]	1.64 (1.23 ,2.18)	2.51 (1.68 ,3.75)
(63,70]	1.81 (1.3 ,2.51)	2.27 (1.43 ,3.59)
(70,77]	1.64 (1.12 ,2.41)	2.22 (1.35 ,3.66)
(77,84]	1.78 (1.15 ,2.77)	2.45 (1.44 ,4.17)
(84,91]	1.36 (0.81 ,2.3)	1.94 (1.06 ,3.54)
(91,98]	0.69 (0.37 ,1.3)	1.65 (0.86 ,3.14)
(98,105]	0.91 (0.46 ,1.79)	2.13 (1.04 ,4.35)
(105,112]	1.09 (0.58 ,2.06)	1.78 (0.84 ,3.78)
(112,119]	0.69 (0.34 ,1.38)	1.05 (0.45 ,2.44)
(119,126]	1.12 (0.47 ,2.66)	2.29 (0.86 ,6.11)
(126,133]	0.79 (0.33 ,1.9)	1.77 (0.65 ,4.79)
(133,140]	0.71 (0.24 ,2.1)	1.29 (0.39 ,4.34)
(140,147]	2.63 (0.62 ,11.08)	2.01 (0.38 ,10.69)

Discussion

Our study sought to explore the association between different strategies for the implementation of patient indexing. We defined as effective indexing, the registration of patients who seek to effectively attend the health center for medical reasons. The analysis of the results of our model shows evidence in favor of the High and Low intensity interventions, at least during the first 90 days. During this period, High and Low intensity interventions were more frequently associated with effective indexing. It is possible to think that the Intermediate intensity intervention, which focused on indexing people who attended the health center mostly for non-medical reasons, does not lead to an effective indexing, since these patients do not concur with the objective of receiving care. After 90 days, all three strategies were equally effective. These results can be explained in different ways. On the one hand, it would be expected that over time all registered patients, sooner or later, will visit the clinic, independently from which indexing strategy they were exposed to. On the other hand, it could respond to intrinsic and subjective factors of the patients that we are not capturing in our model, so there would be residual confusion that would not allow us to see differences between strategies. Finally, the typology constructed does not represent a static model associated with each health center, but is subject to modifications over time. The centers can modify their strategy to political, managerial and technical factors.

It was striking to observe the relatively greater effectiveness of the Low intensity strategy compared to the Intermediate intensity strategy. The centers that adopted Low intensity strategies mainly focused on registering patients during certain hours. This, in many cases, was due to a policy of the center's management to accompany the overall EHR implementation

process, but without complete commitment. This highlights one of the biggest, if not the biggest, problems we experienced during the whole process which was the lack of governance. Even though all centers are part of the same public health system, they act independently. Even within each center, governance is a problem, since in many of them the medical management does not have governance over the administration staff or even the medical staff.

In contrast, many of the Intermediate indexing intensity centers chose to promote indexing in instances not related to health care. Therefore, the majority of the registered people did not correspond to people in need of being seen by somebody from the health care staff in the short term. This, coupled with a partial adherence to the use of the EHR by professionals, can explain the difference between both strategies.

Regarding the High intensity strategy, it basically consists of promoting indexing in all the instances (appointments and maternal formula handout or social security realted consultations) coupled with a strong motivation for the use of the EHR, which in many cases, started at the beginning of implementation process. However, even without a formal test, there did not seem to be significant differences between the High and Low intensity strategies. One possible explanation for this may be that the High intensity strategy represents a combination of the Low intensity (i.e., focused on some patients seeking attention in certain time slots) and the Intermediate intensity (i.e., focused on people who attended the center mostly without needing to be seen by the health care staff in the short term).

Like any observational study, our study has limitations, mainly related to potential residual confounding and misclassification, both at baseline and over time. We were not able to collect information regarding characteristics of the center's management, the attitude of the professionals towards the EHR implementation process, and the availability of appointments at the center. On the population side, we could not record data related to the level of disease burden per patient. Also, since we are in the first stage of the implementation of an EHR system we could not assess the impact of indexing on any health care outcomes. As mentioned above, classification error is also a potential source of bias in our estimates. We have confidence in the basic characterization of the centers, however, we believe that there are errors in the classification of centers that occur over time because many, due to internal or external motivations (i.e., implementation in other centers, direct intervention of the ministry of health), changed the intensity of the strategy initially chosen.

Conclusion

As part of the project to implement an EHR system in the city of Buenos Aires, there are still more than 25 centers to be implemented in the next year. The results of our study provide us with better information that can be used when discussing and negotiating with the management of each center. This includes the possible benefits of each strategy when selecting the one that best suits the needs of the center and the overall implementation process of the EHR.

References

- [1] J. Zech, G. Husk, T. Moore, and J.S. Shapiro, Measuring the Degree of Unmatched Patient Records in a Health Information Exchange Using Exact Matching, *Appl Clin Inform* 7 (2016), 330-340.

- [2] B.H. Just, D. Marc, M. Munns, and R. Sandefer, Why Patient Matching Is a Challenge: Research on Master Patient Index (MPI) Data Discrepancies in Key Identifying Fields, *Perspect Health Inf Manag* 13 (2016), 1e.
- [3] J. Duncan, W. Xu, S.P. Narus, S. Clyde, B. Nangle, S. Thornton, and J. Facelli, A focus area maturity model for a statewide master person index, *Online J Public Health Inform* 5 (2013), 210.
- [4] M. Lee, E. Heo, H. Lim, J.Y. Lee, S. Weon, H. Chae, H. Hwang, and S. Yoo, Developing a common health information exchange platform to implement a nationwide health information network in South Korea, *Healthc Inform Res* 21 (2015), 21-29.
- [5] J.K. Rotich, T.J. Hannan, F.E. Smith, J. Bii, W.W. Odero, N. Vu, B.W. Mamlin, J.J. Mamlin, R.M. Einterz, and W.M. Tierney, Installing and implementing a computer-based patient record system in sub-Saharan Africa: the Mosoriot Medical Record System, *J Am Med Inform Assoc* 10 (2003), 295-303.
- [6] M. Giussi Bordoni, F. Plazzotta, A. Baum, C. Ilc, and F. Gonzalez Bernaldo de Quiros, Elaboración e implementación de una Agenda Digital en Atención Primaria en la Ciudad Autónoma de Buenos Aires, in: *15th Brazilian Health Informatics Conference*, Goiânia, Goiás, Brazil, 2016.
- [7] M. Hammersley and P. Atkinson, *Ethnography, principles in practice.*, Routledge, London, 1989.
- [8] C. Robson, *Real world research. A resource for social scientists and practioner researchers*, Blackwell, Oxford, 2002.
- [9] S.J. Taylor and R. Bodgan, *Introduction to Qualitative Research Methods*, Wiley, 1998.
- [10] B.G. Glaser and A.L. Strauss, *The discovery of grounded theory*, Aldine Publishing Company, New York, 1967.
- [11] P. López-Roldán, La construcción de tipologías: metodología de análisis en Papers, *Revista de Sociología* 48 (1996), 20.
- [12] J.C. McKinney, *Tipología constructiva y teoría social*, Amorrortu, Buenos Aires, 1968.

Address for correspondence

Dr. Santiago Esteban.

Email : santiagoesteban@gmail.com

Design and Construction for Community Health Service Precision Fund Appropriation System Based on Performance Management

Xing Gao, Yao He, Hongpu Hu

Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China

Abstract

Allowing for the differences in economy development, informatization degree and characteristic of population served and so on among different community health service organizations, community health service precision fund appropriation system based on performance management is designed, which can provide support for the government to appropriate financial funds scientifically and rationally for primary care. The system has the characteristic of flexibility and practicability, in which there are five subsystems including data acquisition, parameter setting, fund appropriation, statistical analysis system and user management.

Keywords:

Financial Support; Information Systems; Community Health Services

Introduction

Nowadays, some problems exist in health service in China, that is, higher quality resource are mostly concentrated in metropolitan and big hospitals, and capacity of primary health service is low, hardly playing the role of “health gatekeeper” [1-2]. In order to improve the situation, China has proposed a series of health policies to guide the quality of health resources to sink at the grass-roots level such as “hierarchical medical system” and “the equalization of basic public health services”, which makes positive effects to promote the overall service ability of primary health care. Appropriate economic incentives help improve quality of care and accelerate the policies into effect [3]. For example, the United Kingdom, as early as in 2004, carried out pay-for-performance programs in family practices to improve primary care performance by financial incentives. The payments are in addition to the practices’ core funding, which is based on the number of patients, adjusted for characteristics of the patients and the area [4]. In the United States physicians are paid according to the number of population serviced and completion of performance targets. Meanwhile, additional awards are given for those with outstanding performance [5]. Unlike in Europe and America where practitioners or physicians are the object of funding [6], in China the community health service organization is regarded as the operational unit of primary health care. District governments are in charge of fund allocation for these organizations. Although it seems fair that central government spends funds to every public primary health organizations by the standard of about 45 yuan per resident per year (according to the standard of 2016) to safeguard residents’ basic health, there exist differences in the characters of population serviced and degree of economic development among these organizations, even in the same district. For example,

organization A which serves for 10000 residents can receive 450 thousand, while B serves for 20000 residents can receive 900 thousand. Most residents in A are the old, children and pregnant women which bring out heavy workload to the medical staff in A, while most in B are young person who have a little demand for basic public health service. The more work to do, the less to earn. In this situation, it is not reasonable to allocate fund just by the number of population within respective jurisdictions. Therefore, it is necessary to design and construct a community health service precision fund appropriation system based on the performance management. The so-called precision refers to quantitative and refinement of the indicators related to funding appropriation [7]. In this system, the funds to be appropriated will be scientifically calculated and the change trend of key indicators of basic public health service will be analyzed.

Methods

Demand Analysis is applied to design community health service precision fund appropriation system, including business process analysis, data process analysis and function design.

Business Process Analysis

Business process analysis is a process to sort out and refine the abstract business process of the system. Through analysis, you can clearly understand the business activities and the corresponding responsibilities of the various participants in each business link. The business flow chart is an important tool for business process analysis which describes business dealing process through some provided symbols and lines, to help developers easily and quickly understand the specific business process. Based on the optimization of the original business processes, this paper draws the business flow chart of the precision fund appropriation system among relevant subjects.

Data Flow Analysis

This analysis is helpful for developers to understand the data flow including data generation, data processing and data storage, providing favorable conditions for data structure construction and function module design of the system. Data flow analysis is mostly realized through the data flow chart [8]. We use data flow analysis to clear the data process procedure including data collection, data integration and data utilization in fund allocation of primary care.

Function Design

Based on business process analysis and data flow analysis, the function design should be made in order to realize targets of system. The Function of system is constituted by some subsystems whose interfaces should be clear and functions be

independent of each other. In this paper, the subsystems include data acquisition, parameter setting, fund appropriation, statistics analysis, user management. In addition, in order to increase the practicability and flexibility of the systems, users can choose data collection method, adjust indicator system according to actual demand, which will be specifically introduced in the subsequent results.

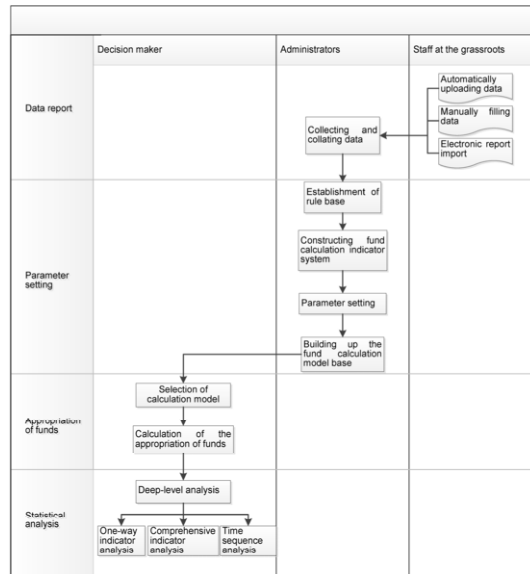


Figure 1 – Business Processes of Precision Fund Appropriation System

Results

The business process analysis for fund allocation in primary care mainly involves three participants and four business links. First, in the data report link, staff of the community health service institutions is in charge of uploading data to the precision fund appropriation system. The uploading method includes automatically uploading and manually uploading. The system administrator is responsible for classifying and sorting out the data. In the parameters setting link, any setting and maintenance is done by the system administrator. This link is the core of the whole business process, but also the foundation for the precision appropriation of funds. There are three main tasks for the system administrator in this link. First, to revise and perfect the rule base, which is used to store the standardization method of each type of business data; it needs to be updated and maintained regularly. Second, to maintain the fund appropriation indicator system template and modify the indicators or the weights of the template to form a new fund appropriate indicator system. Third, to regularly summarize the fund appropriation indicator system, modify the parameters setting and establish fund appropriation model base according to user needs. In the fund appropriation link, the decision makers, according to health strategy development priorities, select the appropriate fund appropriation model to calculate the amount of the funds to be appropriated to the community health service institutions within its jurisdiction. The statistical analysis link is used for deep-level mining and analysis of fund appropriation data. In this link, decision makers can understand fund appropriation statuses, operational and developmental situations of different

communities from different perspectives (as shown in Figure 1).

Data Flow Analysis

The precision fund appropriation data flow will include data classification, standardized processing, appropriation indicator system's weight settings, fund calculation (as shown in Figure 2).

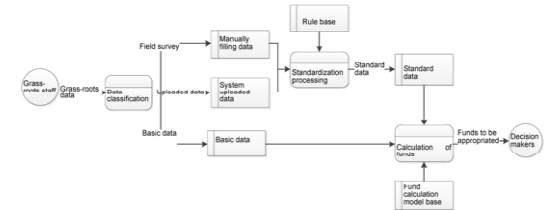


Figure 2 – Precision Fund Appropriation System's Data Flow Chart

The sources of the original data of precision fund appropriation are from the community health service institutions. First the data will be classified according to the data types and characteristics after being uploaded. Data type includes uploaded data, field survey data and basic data. Among them, uploaded data is regarded as the basic public health service data uploaded directly from the information system, such as electronic health record filing rate, neonatal follow-up rate. Unlike the uploaded data, the field survey data is obtained only from field investigation rather than automatic acquisition from information systems, such as health education plan and summary, and health education activities. The basic data refers to the community health service institutions' service scope, population characteristics and other attributes, such as important crowd proportion, population density. Second, data standardization will be conducted. That is, choose different rules from the rule base for grass-roots data standardization processing. By defining standard workload and using comparison by flexibility, standard time or service process, the administrator set the corresponding workload value of each indicator to complete data standardization and obtain standard data. Third, call the fund appropriation indicator system and set the weight of each indicator, and then multiply the standard data of each indicator by the weight of each indicator, and use their multiplication results as the total workload values of the community health service institutions. Last, after the annual basic public health service funds are confirmed, the amount of the funds which should be allocated to the primary health institutions will be provided to decision makers.

Function Design

Based on the business process analysis and data flow analysis, a community health service's performance management fund appropriation system is designed. There are five functions including data acquisition, parameter setting, fund appropriation, statistics analysis, user management and etc..

Data Acquisition

The data acquisition module is used to collect and collate the business data and basic data associated with the fund appropriation. This module supports three kinds of data acquisition modes such as online data acquisition, local data uploading and software interface data import. Other non-online data can be manually put in or imported by other software through the system interface.

Parameter Setting

The parameter setting module is used for setting attributes of the business indicators related the fund appropriation, including

rules base setting, fund appropriation model base settings. The rules base is used for storage of a series of rules of business data standardization processing, such as automatic scoring rules, workload standardization rules. The fund appropriation model base is provided for setting and updating fund appropriation indicator system such as indicator increase, deletion, weight setting. Moreover, through free grouping of relevant business indicators, a new fund appropriation indicator system can be formed to meet the needs.

Currently, the indicator system for fund allocation used in the system is based on the national basic public health service item specification [9-10]. The indicator system is divided into three levels. In level one, there are three indicators, including public health services, population characteristics and regional characteristics. In level two there are 16 indicators, mainly including electronic health records, health education, preventive inoculation, population density and terrain features. In level three, there are 36 indicators, mainly including electronic health records filing rate, health education activities, early pregnancy rate, elderly health management rate, rate of health management of patients with diabetes, the elderly coefficient.

Fund Appropriation

The fund appropriation module, by selecting the appropriate fund appropriation indicator system, automatically calculates the amount of financial funds that governments should appropriate to the community health service institutions within their respective jurisdictions. The module includes calculation method choice, results inquiry and other functions. The calculation method selection refers to selection of the suitable appropriation models in a variety of fund appropriation models, including fee for service, work quality priority, assessment results priority and other models. Calculation results query is used for viewing and comparing the amount of funds appropriated to the primary health institutions in a same district. In addition, the module can also call the fund allocation model base, and, according to the needs modify the indicators and weights in the indicator system.

Statistical Analysis

The statistical analysis module can be used to view the ranking of the key indicators of all institutions according to the development direction of the primary health services. For example, in-depth analysis can be conducted from the quantity, quality, level of technological innovation and progress and other levels, so as to provide reference basis for the provision of special awards. At the same time, the module should be used to query the historic fund appropriation indicators and related data, carry out comparison analysis graphically with the historic fund appropriation indicators, calculate input-output ratio and the change rate of fund appropriation for each institution, so that users can combine the history and current reality to make minor adjustments to the fund appropriation parameters of the current year.

User Management

The user management module is used for management of adding, deleting system users and of their permissions. Permission management refers to the permissions set for different users to use or access the system resources (function menu items, buttons, input controls, etc.).

Discussion

The results are discussed in context of present and other The UK's pay for performance scheme is similar to the US, they are motivating family doctors to increase the quantity and quality of primary health care by linking the performance score to the

bonus with a strong database. In the UK, the family sign contract with their doctor on a voluntary basis, to enhance service capabilities of family doctors. There are 146 performance indicators for the pay for performance scheme, involving chronic diseases, health care organizations, and patient experience. The indicators are not static and they will be modified, added or removed based on grass-roots services [1]. At the same time, each indicator is assigned a different score according to the degree of difficulty. The performance score of the family doctor is determined by the degree of completion of the indicators. The final performance bonus is the product of the unit score and the performance score. The data required for performance bonus calculations can be extracted directly from Quality Management and Analysis System (QMAS). Besides, there is an exceptional reporting which is applied to patients who have coexisting conditions that affect their optimal care. Family practitioners can also increase their income by inappropriately excluding patients for whom they have missed the targets. The UK's basic health services funding is also similar to the US, the amount of payment is based on the number of service groups, the degree of completion of performance and unit standards. At the same time, US CAHPS database worked with Microsoft to provide evidence-based support and results feedback for developing performance management standards through the collection and storage of national performance management data, comparative analysis of different regions and periods of performance benchmarks.

In order to enhance the grass-roots service capacity, China is also constantly exploring the relationship between community health service performance management and funding allocation. China's financing model, allocation methods, service groups, data collection are different from Europe and the US because of the national conditions, economic development and the degree of development of information technology.

In the financing model, China's primary health care services are central, provincial, municipal and district finance co-ordination. Due to regional economic differences, funding for primary health care services varies from region to region. Take basic public health services as an example, the fund of basic public health service per person per year is 73 yuan in Nanjing, Jiangsu, while Hefei, Anhui is 45 yuan in 2016. The reason for the difference is because the economy is more developed in Nanjing, Jiangsu Province, so in addition to the central government allocated 45 yuan, as well as from the provincial, municipal and district-level financial support 28 yuan. But in Hefei, some areas follow the example of Europe and the United States to carry out the way according to the project, and some areas based on the existence of "eat big pot" situation.

In the allocation of funds, Europe and the United States allocated to the object is a family doctor, in accordance with the service and service difficulty to reward, the upper limit of funding is no special restrictions. The main body of primary health care services in China is the grassroots health institutions, the district-level government allocate funds to the grassroots health institutions. Due to limited economic level and the total amount of funding is fixed, it is necessary through the performance management of the funds reasonable distribution.

In the service groups, China's primary health care institutions are pushing the "grass-roots first diagnosis, two-way referral" strategy. But the current form of service is for the elderly, pregnant women and children to provide basic medical and basic public health services. This is the gap between Europe and the United States on the grassroots "health gatekeepers" position.

In the data collection, unlike the UK to summarize, integrate and utilize national primary health data by QMAS, as China's system construction standards are not unified, the current grassroots health data in China only to achieve the district-level sharing, while interconnection between grassroots systems needs to be improved. Therefore, we collect data in combination with automatic and manual combination when we allocate funds.

Due to the difference of characteristics of the service population, the degree of economic development and the data sharing, we designed the Precision Fund Appropriation System Based on Performance Management based on performance management based on the principle of flexibility and intelligence. The main features are as follows: (1) There are a variety of funding estimate indices, district-level government can be combined with the actual needs of the corresponding index system, and can be added to the indicators, delete and weight configuration; (2) The quantitative index can be automatically scored according to the pre-set calculation rules, which can reduce the workload of staff to manually score. Take the example of electronic health file index, if the file rate is greater than or equal to 70% had 10 points, if less than 70% will not score. (3) Taking into account the construction of grassroots information platform, in the data acquisition both automatic acquisition function and manual upload function, which are used to compensate for the low degree of data sharing defects.

Conclusion

Allowing for the differences in economy development, informatization degree and population characteristic served among different districts, the system designs functions with a certain flexibility and practicability, which can provide support for the government to appropriate financial funds scientifically and rationally for primary care.

Acknowledgements

This work is supported by the national natural science foundation of China (71273280), and special fund for basic scientific research business of central public research institutes (2016RC330013).

References

- [1] Duan Ting, Gao Guangying, Wang Jing, Li Xin, Jia Jirong, Zhu Junli. Empirical Study on Community Public Health Cost Calculation and Fund Compensation: a Case Study of Fangshan District, Beijing, *China Health Economics*, **1** (2015).
- [2] Gao Xing, Chen Quan, Lei Xingyun, et al. Study on the Status Quo & Problems and Countermeasures of China Community Health Service Performance Management, *China General Practice*, **16** (2014), 1826-1828.
- [3] Campbell S M, Reeves D, Kontopantelis E, et al. Effects of pay for performance on the quality of primary care in England, *New England Journal of Medicine*, **361** (2009), 368-378.
- [4] Doran T, M. P H, Fullwood C, et al. Pay-for-performance programs in family practices in the United Kingdom, *New England Journal of Medicine*, **355** (2006), 1832-1833.
- [5] Rosenthal M B. Early Experience With Pay-for-Performance: From Concept to Practice, *Journal of the American Medical Association*, **294** (2005), 1788-93.
- [6] Glanville J, Kendrick T, McNally R, et al. Research output on primary care in Australia, Canada, Germany, the Netherlands, the United Kingdom, and the United States: bibliometric analysis, *BMJ* **342** (2011), d1028.
- [7] Liu Tianwei, Zhang Yongjia, Zhu Huirong. Discussion on the Application of Project Management Informationization in the Refinement Management of Community Health Service Center, *Chinese Journal of Medicinal Guide*, **17** (2015), 431-432.
- [8] Xue Huacheng. Management Information System. The Sixth Edition, *Tsinghua University Press*, 2013
- [9] Ministry of Health. 2015 National Basic Public Health Services List. Retrieved from <http://www.nhfp.gov.cn/jws/s3577/201506/61340494c00e4ae4bca0ad8411a724a9.shtml> 2008[EB/OL].
- [10] Liu Bao, Hu Shanlian, Xu Haixia, et al. Study on Equalization Indicator System of Basic Public Health Services. *Research on China's Health Policy*, **2** (2009) 13-17

Address for correspondence

Correspondence: Dr. Hong Pu Hu,
Address: Chinese Academy of Medical Sciences, Institute of medical information, No.3 Yabao Road, Chaoyang District, Beijing, 100020, PR China
Tel: 010-52328870
e-mail: hhp4430140@sina.com.

Ethical Issues in Implementing National-Level Health Data Warehouses in Developing Countries

Milka B. Gesicho^a, Troy D. Moon^b, Elizabeth Heitman^c, Martin C. Were^{a,b,d}

^a Institute of Biomedical Informatics, Moi University, Eldoret, Kenya,

^b Vanderbilt University Medical Center, Vanderbilt Institute for Global Health, Nashville, TN, USA,

^c Program in Ethics in Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX, USA,

^d Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA.

Abstract

National Health Data Warehouses (NHDWHs) promise to improve individual and population health. These systems are now increasingly being deployed in low- and middle-income countries (LMICs), given increased recognition of the value of data analytics in informing decision-making. However, ethical issues relevant to implementation of NHDWHs in LMICs remain largely ignored or, at best, inadequately addressed. In this paper, we highlight and critically analyze several of the key ethical issues, including privacy, informed consent and trust, confidentiality and security, secondary data use, sustainability and implementation validity, risk-benefit ratio, governance and conflict of interest, justice, equity, access, and collaborative partnerships. We then provide a set of guiding principles and points to consider for countries and implementers to guide further decision-making around ethics of implementing NHDWHs within LMICs.

Keywords:

Ethics; Information Systems; Developing Countries

Introduction

A national health data warehouse (NHDWH) is an extract of health-related information from multiple organizations and data sources within a country, which is used primarily to facilitate decision-making through various data analytics approaches.[4] NHDWHs can help in (a) strategic planning and quality improvement initiatives at the organizational, state, and national levels, (b) personalization of patient-level care services to improve clinical outcomes, and (c) support for public health and surveillance initiatives, among other activities. In the age of personalized medicine, NHDWHs are becoming key to ensuring the right treatment is provided to the right person at the right time.

Within the last few years, an increasing number of low- and middle-income countries (LMICs) have developed NHDWHs. As an example, Kenya, through its Ministry of Health (MoH) and the National AIDS and STI Control Program, has worked with international implementing partners to create a NHDWH for HIV programs. The initial aim of Kenya's NHDWH is to incorporate all patient-level HIV data from key digital health systems from across the country. In Zambia, there was a request in 2016 for proposals for the 'Data Warehouse for HIV/AIDS Policy and Program Decision-making in Zambia under the President's Emergency Plan for AIDS Relief (PEPFAR)' program.[2] These and many other examples indicate that the use of NHDWHs within LMICs will only gain increased prominence in practice. As is typical in many LMIC settings, HIV programs have taken the

lead, but undoubtedly, NHDWHs will become relevant for multiple other purposes in the future.

The NHDWHs being implemented currently contain information with various levels of granularity, including patient-level data. Despite growing enthusiasm by countries and funding organizations to implement NHDWHs, very little attention has been paid to the ethical issues that surround their adoption. In fact, while the benefits of NHDWHs might look obvious, few have stopped to ask the numerous pertinent ethical questions surrounding their use in LMICs. These ethical questions are relevant at every stage of the data warehouse, including the extraction of data from primary sources, data transmission procedures (especially in settings with limited infrastructure), the creation of "data marts" where data are curated to suit particular needs, and the data analysis and dissemination stages. Countries need guidance on a systematic approach for identifying and evaluating ethical issues that arise when NHDWHs are being considered. While ethical issues in LMICs may well be the same as those in developed countries, some variability will exist given several unique considerations within LMICs. Among the differences that could impact ethical consideration are the levels of stigma associated with diseases such as HIV, poorly-defined policies to protect patient rights, central and often paternalistic mechanism for decision-making, variable literacy levels, and variable resources. Nonetheless, an examination of ethical issues surrounding NHDWHs within LMICs will still be instructive and applicable for developed countries. To date, no comprehensive and rigorous analysis exists to provide ethical guidance around NHDWHs for LMICs. In this article, we critically explore the range of ethical issues that can arise from implementation and use of NHDWHs in LMICs. For this work, we take a systematic approach guided by widely accepted ethical frameworks. We then propose a set of 'Points to Consider' to guide further thinking and decision-making.

Approach

Components of a NHDWH

A data warehouse consists of several components that are relevant to assuring the appropriate integration of data from multiple source systems, segmentation of these data for various uses, and reformatting to allow efficient implementation of analytic procedures (Figure 1).

Primary Data Sources:

Primary data sources are used to capture digital data mostly at the point of service. Common primary data sources for health data warehouses include electronic health record systems, pharmacy information systems, laboratory information

systems, health management information systems, logistics management information systems, and finance systems. Data from source systems are usually extracted, transformed, and loaded (ETL) into data warehouses using various mechanisms.

Data Warehouse:

The data warehouse is a collection of data gathered from one or more primary operational data systems and organized in a way that optimizes analysis and projection. In most cases, the data in the warehouse are subject-oriented, normalized, and have data/time identification to provide a snapshot at any given time.

Data Mart:

Often a subset of the data warehouse, a data mart represents a subject-specific extraction of a portion of the data in the

warehouse, organized in a format that allows particular types of analyses and efficient querying. In NHDWHs, examples of data marts can include patient encounters, drugs, and laboratory-based data.

End-user Tools:

End-user tools include a range of applications to facilitate querying of the data, report writing, and a wide array of analytics. End-user tools also provide relevant interfaces and dashboards for a unified view of analytics and performance data needed. Advanced business intelligence tools now exist to facilitate data mining, statistical analyses, and reporting.

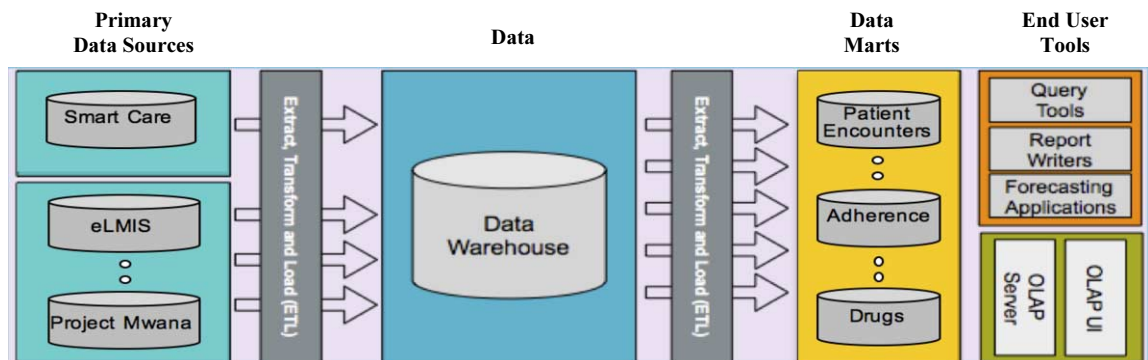


Figure 1 - Schematic of the various stages in NHDWH using systems in Zambia as an example. Figure adapted from presentation by Dr. Dipankar Sengupta. Introduction to Knowledge Discovery Process. Mar 21, 2016. Eldoret, Kenya.

Ethical Framework

A rigorous analysis of ethical issues surrounding NHDWHs in LMICs requires a systematic approach. Several existing works can be adapted to inform a comprehensive analysis of the ethical issues surrounding NHDWHs in LMICs. Among these are: (a) ethical guidance around evaluating clinical research in developing countries;[7] (b) ethical frameworks around implementation of electronic record systems in developing countries;[10] and (c) various research highlighting ethical issues in big data analytics.[5; 6] Using a combination of these approaches, we prioritized several key ethical issues that are further outlined below.

Ethical Issues in NHDWHs in LMICs

Privacy, Informed Consent, and Trust

A key component of privacy is the control an individual has over the extent, timing, and circumstances of sharing information about themselves to others. Coordinators of NHDWHs often do not seek consent for the inclusion of individual-level information. Conflicts can easily arise regarding ownership of data within NHDWHs. It is not uncommon for Ministries of Health (MoH) to claim ownership of patient-level data, often insisting that they have no need to seek approval from the individuals to use this data. It could be argued that expectation of full consent on collection, storage, and use of data might be unrealistic, especially in settings where individuals have limited literacy and understanding of digital technologies. Whether or not individual consent is sought, custodians of NHDWHs still have the responsibility to inform individuals about the use of their data in a culturally sensitive format. Some of the information worth sharing with

individuals whose information will be stored in a DWH includes the types of secondary data analytics to be conducted, information on patient rights, data ownership, and potential risks to individuals in having data within the NHDWH. Public and individual expectations should also be appropriately managed.[10]

Individuals and providers may not trust those charged with custody and analysis of data contained within NHDWHs. It is not a stretch to imagine situations in which the stored data could be used to target stigmatized groups, tribes, or individuals, such as individuals with a particular sexual orientation. In many cases, simply assuring people that their data are protected will likely not satisfy their concerns, and may heighten their concerns about privacy. Appropriate controls and clear disclosure are, therefore, needed to ensure that custodians of the data do not abuse this trust.

Confidentiality and Security

The ethical principle of confidentiality dictates that information gathered about individuals is not revealed without appropriate authorization and that it is securely maintained. Confidentiality is often best assured when information that is not required is never collected at all. At the very least, limits must be placed on the amount of individually identifiable information that is collected.

In LMICs, NHDWHs pose a risk to confidentiality as clear guidelines often do not exist to limit amounts and types of information extracted from primary sources. Often, regulations that restrict the use or disclosure of Protected Health Information (PHI), akin to the Health Insurance Portability and Accountability Act (HIPAA) in the United States, do not exist in these countries.[3] Even when primary data sources do not collect PHI or when information is 'de-

identified' prior to incorporation into the NHDWH, the simple act of aggregating data from various disparate sources into one DWH increases the chance of re-identification as this information can sometimes be combined in novel ways.

NHDWHs are an attractive target for data breaches given the sheer volume of data contained. With sophisticated attacks on some of the world's largest software companies and healthcare organizations, it should be no surprise that NHDWHs in LMICs will be targeted. Beyond threats of breaches from outside the DWH, particular attention also needs to be paid to threats from within, given the potential lack of systematic, rigorous, and continued training of all personnel with access to the DWH and laxity in enforcing existing controls. The mechanisms used to transport or transmit data in LMICs may also increase security vulnerabilities. Where there is inadequate network infrastructure, insecure 'sneaker networks' (such as use of flash or external hard drives) are often employed to transfer data into the NHDWH. The use of local servers for the DWH, as is often dictated by national policies that restrict cloud-based storage, also poses a significant physical security threat.

Secondary Data Use

Secondary use of data is closely connected to several other ethical issues, particularly privacy, trust, and confidentiality. By nature, analytics conducted from within the NHDWH uses data that was initially acquired for a different purpose. To avoid unauthorized disclosures and related harm to individuals and groups, as well as their mistrust of the DWH itself, constraints must be placed on the types of analytics that are allowable. Where appropriate, data should be de-identified prior to aggregation into the DWH, and resulting analyses should not allow re-identification unless re-identification will explicitly benefit the individual or society and the process is disclosed to those individuals whose data are re-identified. Additionally, the methods and location of access to NHDWHs by secondary users ideally need to be controlled. It would also be wise for countries developing NHDWHs to constitute oversight bodies tasked with specifically reviewing the appropriateness of proposals for secondary data analysis. While the benefits of secondary analysis may be significant, care must be taken to ensure that these data are not used for mass population surveillance, to target individuals or populations, nor to restrict human and political rights.

System and Implementation Validity

Many countries are advocating for NHDWHs without the appropriate standards, guidelines, and enforcement mechanisms being in place to guide implementation of electronic databases generally. Often overlooked are critical assessments of the suitability and appropriateness of implementing NHDWHs in settings that have significant barriers around capacity, cost, and infrastructure. Best practices for data warehousing do exist, and it is recommended that the implementation of NHDWHs be gauged against these standards.[9]

Data quality presents a major challenge in LMICs where medical recordkeeping may be inconsistent and documentation incomplete. Quality can be compromised from the primary source, during the data extraction and transformation stages, and during aggregation for analysis. The saying 'garbage in garbage out' certainly rings true in NHDWHs with quality issues arising around completeness, accuracy, timeliness, uniqueness/duplication, validity, and consistency of the data. At the same time, the development and maintenance of NHDWHs offers a unique opportunity to uncover key gaps in data and other quality issues within the

care system. Establishing and operating a NHDWH would permit systematic improvement of these shortcomings through approaches such as patient matching and data de-duplication that would not be possible where the data are not warehoused.

Favorable Risk-Benefit Ratio

When used with appropriate analytic tools, NHDWHs can provide business intelligence with the potential to improve individual health outcomes, and the efficiency, equity, and governance of health systems. At the same time, aggregating data in one place and conducting various types of analytics may present numerous risks. For these reasons, there should be a clear determination of who is expected to benefit from the implementation of the NHDWH and who will bear the risks. NHDWHs and business intelligence should have a reasonable probability of a favorable risk-benefit analysis for the individuals, with a greater social value to the community and country to justify implementation.

In an ideal world, the determination of the risks of harm and potential benefit should be explicit, objective, and based on well-outlined criteria. However, these valuations tend to have a significant subjective component, often shaped by societal and political values. There is a risk of overstating the benefits of NHDWHs as most analytics could be done with single-source systems, without the need for aggregation. Where using more data looks like an advantage, despite the dearth of evidence on the benefits and risks of NHDWHs in LMICs, countries often proceed with NHDWHs based on inadequate information.

The implementation of an enterprise-level NHDWH comes with significant initial and ongoing costs, up to millions of dollars, especially when they use market-leading but proprietary data warehousing systems like Oracle or Teradata.[1] Use of open-source systems should be considered to help reduce these costs. In many instances, countries will not have adequately considered the long-term sustainability of NHDWHs, with primary funding for their development and implementation coming from time-limited grants. An opportunity exists to implement self-sustaining business models for NHDWHs, given that there are multiple stakeholders, such as insurance and pharmaceutical companies, who have the ability to pay for knowledge derived from NHDWH.

Where resources are constrained, resource allocation among multiple competing priorities is an essential consideration. Even when the cost-benefit of a NHDWH is evident, the decision whether to allocate resources for its implementation has to be evaluated against using those resources on other proven interventions, such as immunizations or safe water. Through this comprehensive resource-allocation lens, a NHDWH might end up being lower on the priority list.

Governance and Conflict of Interest

Good governance at multiple levels is essential for success of any NHDWH project. The primary implementing team needs to be well qualified and managed and stakeholders appropriately involved. A multi-stakeholder technical working group should guide implementation, while an independent body should be engaged for ethical review. Significant financial oversight is also necessary, given the large amounts of money often involved in these projects. Accountability and transparency thus go hand in hand.

NHDWHs present multiple potential areas for conflict of interest. The MoH, often the primary custodian of the data, might be resistant to data analytic practices that uncover deficiencies within the system or that might portray the

Ministry or other government agencies in a bad light. Further, as an arm of the state, there might be undue pressure to use the data within the NHDWH for politically motivated surveillance. NHDWHs, also offer a significant opportunity to make money, raising in turn the potential for financial conflicts of interest, especially when those charged with making financial decisions work within a culture that may struggle with corruption. Monitoring by regulatory bodies and other oversight systems are, therefore, needed to ensure that both real and perceived conflicts of interest are avoided.

Justice, Equity and Access

Through data analytics, NHDWHs can be used intentionally or unintentionally to reveal injustices in the distribution of resources involving personnel, finances, supplies, equipment, and health care facilities. In LMICs, where unequal access to resources is often a significant ethical challenge, NHDWHs can be an important tool in achieving distributive justice and health equity. NHDWHs can be appropriately used to identify such issues as the prevalence of domestic violence, untreated marginalized populations, geographic health disparities, and abusive practices such as female genital mutilation, in order to address them. However, the quest to discover distributive injustices through NHDWHs should be deliberate, and coordinators of the DWH need a plan for using this information to ensure improved access to quality care services for all.

Collaborative Partnerships

Multiple stakeholders are directly affected by or involved with the implementation of NHDWHs, ranging from individuals and communities whose data are placed into the warehouse, the MoH and other government bodies, technical support and implementing organizations, researchers and funders. Collaborative partnership requires that these stakeholders be appropriately engaged in ethical discussion, especially regarding the aspects of the NHDWHs that have direct impact on them, before, during, and after implementation. True collaborative partnership requires adequate and appropriate representation of each group. Community representatives and leaders can be entrusted to act on behalf of individuals, even when these individuals might not completely comprehend the various dimensions of the NHDWHs and its associated knowledge discovery processes.

Defining how the data within the NHDWH is used also must be a shared responsibility, to ensure that priorities and concerns of all parties are adequately represented. There is always a temptation of professional paternalism around the use and management of data, with a risk of exploitation by the custodians of the data. In essence, analytics should be done *‘with people and not on people’*. Use of consultative forums and groups, and assembling of working groups on various questions can help to maintain mutually respectful participation by relevant stakeholders, with all voices listened to and heard appropriately. This type of engagement is key to the sustainability of the NHDWH and ongoing trust in its greater purpose and management.

Guidance and Points to Consider on NHDWHs

It is evident that a range of ethical issues arises with the growth of NHDWHs in LMICs. MoH and key stakeholders must consider and be equipped with ethical guidance on how to approach these issues, with specific attention to differences between LMICs and industrialized countries where DWHs are now common. Below, we provide a set of points to consider that form a basis for guidance on how to approach some of

these ethical issues. However, as noted by Fife et al. about their recommendations for faculty ownership of medical facilities, these points “are neither a set of decision rules that mechanistically resolve issues at stake, nor are they a set of principles from which interpretation can be manipulated by various parties to support their particular points of view”.[8] This guidance aims to help policy makers reach sound and principled ethical judgments and to begin a conversation about ethically defensible approaches for implementing NHDWHs in LMICs.

Confidentiality and Security

It would be advisable for LMICs to develop their own clear guidelines on how to manage, protect, and secure the large volumes of data contained within their NHDWHs. Identifiable information that has to be collected also has to be explicitly outlined a priori, with justification as to why they are needed. Whenever possible, data stored within the NHDWH should be de-identified and controls placed on procedures that would risk re-identifying PHI. Appropriate technical, training, administrative, and physical security systems and procedures should be part of any implementation, and these should be continuously updated to fit the dynamic environment of health data warehousing.

Privacy, Informed Consent and Trust

Individuals and communities in LMICs might not be very familiar with particular technologies, and different approaches will likely be required to help them understand what it means to have their data as part of the NHDWHs. Culturally-sensitive educational approaches for the population, use of key community representatives, and transparency are key to ensuring that custodians of the data can gain trust from stakeholders. Decisions have to be reached collaboratively on the rights and ownership of the data stored within the NHDWH. In the western world, individuals are typically able to consent to have their data incorporated into a Health Information Exchange, and such a model of consent can be considered at the point of primary data collection.

Secondary Data Use

There should be clarity on how NHDWH data will be used and by whom. An independent body can help to oversee the appropriateness of requests for analytics and access procedures. Identifying potential harms to different populations and communities will be an important role for community engagement, and the principle of first doing no harm (non-maleficence) should always take precedence when dealing with secondary use of data within the warehouse.

System and Implementation Validity

The implementation of NHDWHs needs to be evaluated against internationally accepted best practices for data warehousing generally. Given that most existing national health information standards do not touch on health data warehousing, revisions to these standards are needed to incorporate the multidimensional considerations to data warehousing, with attention also paid to ethical issues. Strict data quality assurance (DQA) procedures should be implemented at each stage of the data warehousing process. Data analytics procedures should also aim to inform improvements on quality and validity of data collected from primary sites into the DWH.

Favorable Risk-Benefit Ratio

Each country needs to assess clearly the financial and social costs of implementing and sustaining a NHDWH. Alternative approaches, such as doing analytics on primary data systems,

should be adequately explored. A NHDWH must be prioritized in the context of other proven interventions, as it will likely compete for limited resources. There should also be clarity on who will bear the costs over time and who will benefit from implementation of the NHDWH. Long-term sustainability plans should be addressed at the outset to avoid huge disruptions that will occur should funding dry up. Finally, a rigorous evaluation agenda, focusing on the benefits and costs of NHDWHs, is needed. Even when the evidence does not exist from the start, continuous re-assessment of costs and benefits are invaluable to justifying continued expense on a NHDWH.

Governance and Conflict of Interest

Appropriate governance structures call for clear assignment of responsibilities for key important tasks related to the DWH. Mechanisms for assigning responsibility and roles have to be above board and attentive to potential conflicts of interest. Relevant oversight should be in place, and an independent body empowered to evaluate appropriateness of implementation and analysis requests. Where significant breaches occur, heavy penalties should be imposed and corrective actions taken to prevent future such events. Proper accounting and procurement procedures are needed, as are independent audits of expenses and activities. Conflicts of interest at organizational and individual levels should be identified early and relevant training provided to all personnel. Transparency and accountability should remain a basic responsibility for every stakeholder.

Justice, Equity and Access

Countries need to work towards appropriate distribution of risks and benefits of data warehousing. Analytics should be used to uncover inequities in access to health services and related resources, and should highlight when one group is favored over others. Care has to be taken not to use the data in NHDWHs to discriminate or to impinge on rights of individuals or communities, even when there are political pressures to do so.

Collaborative Partnerships

Appropriate decisions on whether and how to implement the NHDWH should be reached with input from all stakeholders. Individuals who are vulnerable and whose personal health information will form a core part of the NHDWH should have trusted representatives for them. Concerns and interests of the different stakeholders should be handled and addressed, and the community should be made to feel a sense of ownership of the NHDWH.

Conclusion

Ethical concerns remain largely ignored in ongoing conversations around NHDWHs in LMICs. Without appropriate attention to some of the above ethical principles, implementations of NHDWHs are likely to be plagued with numerous problems. Should the needed attention be paid to these ethical issues, the potential benefits of these NHDWHs will be truly realized, with improved outcome for the individuals, better personalization of their care, and overall improved health of the population.

Acknowledgements

This work was supported in part by the NORHED program (Norad: Project QZA-0484). The content is solely the responsibility of the authors and does not necessarily represent

the official views of the Norwegian Agency for Development Cooperation. All authors report no conflicts of interest.

References

- [1] Bedi H. Understanding the costs of data warehouses. *The Data Center Journal*. February 23, 2015. Available at <https://www.datacenterjournal.com/understanding-costs-data-warehouses/>. Last accessed Dec-18-2016., in.
- [2] Data Warehouse for HIV/AIDS Policy and Program Decision-making in Zambia under the President's Emergency Plan for AIDS Relief (PEPFAR) - CDC-RFA-GH17-1740. Available at <http://open-grants.insidegov.com/1/46080/Data-Warehouse-for-HIV-AIDS-Policy-and-Program-Decision-making-in-Zambia-under-the-President-s-Emergency-Plan-for-AIDS-Relief-PEPFAR-CDC-RFA-GH17-1740>. Last access Dec-18-2016., in.
- [3] Department of Health and Human Services. Standards for Privacy of Individually Identifiable Health Information - 45 CFR Parts 160 and 164. Federal Register. Vol. 65, No. 250. December 28, 2000. Available at <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/privacyrule/prdecember2000all8parts.pdf>. Last accessed Dec-18-2016., in.
- [4] Grob H and Hartzband D. Health Centers and the Data Warehouse. *Health Center Controlled Networks Series, Information Bulletin #14*. National Association of Community Health Centers. September 2008., in.
- [5] Lieke Jetten & Stephen Sharon, Selected Issues Concerning the Ethical Use of Big Data Health Analytics, 72 Wash. & Lee L. Rev. Online 486 (2016), <http://scholarlycommons.law.wlu.edu/wluonline/vol72/iss3/2>, in.
- [6] Martin, K.E., *Ethical Issues in the Big Data Industry*. MIS Quarterly Executive. 2004. 14 (2): p. 67 - 85. , in.
- [7] E.J. Emanuel, D. Wendler, J. Killen, and C. Grady, What makes clinical research in developing countries ethical? The benchmarks of ethical research, *J Infect Dis* **189** (2004), 930-937.
- [8] R.S. Fife, P. Keener, E.M. Meslin, M. Randall, and R.L. Schiffmiller, Faculty ownership of medical facilities: inappropriate conflict or an opportunity that benefits physicians and patients?, *Acad Med* **79** (2004), 1051-1055.
- [9] W.O. Hackl and E. Ammenwerth, SPIRIT: Systematic Planning of Intelligent Reuse of Integrated Clinical Routine Data. A Conceptual Best-practice Framework and Procedure Model, *Methods Inf Med* **55** (2016), 114-124.
- [10] M.C. Were and E.M. Meslin, Ethics of implementing Electronic Health Records in developing countries: points to consider, *AMIA Annu Symp Proc* **2011** (2011), 1499-1505.

Address for correspondence

Martin C. Were, MD MS (martin.c.were@vanderbilt.edu)

Patients' Online Access to Electronic Health Records: Current Status and Experiences from the Implementation in Sweden

Maria Hägglund^a, Isabella Scandurra^b

^aHealth Informatics Centre, Karolinska Institutet, Stockholm, Sweden

^bInformatics, School of Business, Örebro University, Örebro, Sweden

Abstract

The number of eHealth services for patients is rapidly increasing worldwide. This paper describes the status of a very important eHealth service for patients in Sweden, the Patient Accessible Electronic Health Record (PAEHR). As many countries are facing an introduction of national eHealth services providing health information to the patients, lessons learned from Sweden may improve the deployment and use of PAEHRs and similar eHealth services. Challenges that remain in Sweden relate to local differences in the implementation that lead to fragmentation and unequal access to information. Initiatives have been taken to reconcile some of the problems, e.g. an updated national regulatory framework for PAEHR. To date, evaluations are often performed from a healthcare provider perspective, focusing on aspects that are considered important by healthcare professionals and decision makers. Based on experiences of this nation-wide implementation we argue for the need to also base evaluations of eHealth on the perspective of the patients.

Keywords:

Patient Portals; Electronic Health Records; Health Information Exchange

Introduction

eHealth is often suggested to have the potential to revolutionize the way healthcare and prevention is provided, shifting the balance of power and responsibility from healthcare professionals to patients and citizens [1], [2]. Yet, many of the applications developed for patients are either designed from a healthcare providers' perspective, or completely independent from healthcare. Patient involvement in the design of eHealth has been stressed as important to achieve usable and useful eHealth solutions [3]. Experiences of the nation-wide implementation of one of the most important eHealth service for patients in Sweden accentuate the need to also base evaluations of eHealth on the perspective of the patients, rather than, as it today often is, only perform evaluations from a healthcare provider perspective, focusing on aspects that are considered important by healthcare professionals and decision makers.

Patients' online access to EHR

Sweden recently updated the national eHealth vision that now states that all residents from 16 years of age should by 2020 have access to all health related information documented in county-funded health and dental care [4]. However, implementing these eHealth services are controversial for the

healthcare professionals [5] and it is challenging to realize on a national scale [6].

Internationally, there is also a drive towards providing Patient accessible EHRs (PAEHRs), but it has been limited in part by professional resistance and concerns about security and privacy [7][8], legal constraints [9] and low uptake of other online resources for patients. In a systematic review from 2014 [10], a lack of evidence from high-quality studies about the impact of online access was noticed, yet it was clear that the tensions between the growing consumer demand to access data and healthcare systems not yet ready to meet these demands have increased in recent years [10]. Many of the studies identified in the review originated from the USA, from large health plan-based programmes, whereas a minority of studies originated from Europe. The review indicated that patient online access to their EHR and other services offer increased convenience and satisfaction [10]. However, professionals are often concerned about impact on workload and risk to privacy, which is also the case in Sweden [11]. The authors of the review article conclude that a redesign of the business process to engage health professionals in online access and of the EHR may be required to make it easier to use and provide equity of access to a wider group of patients. Another review published in 2015 called for more empirical testing regarding the effect of PAEHRs on health outcomes for patients and healthcare providers [12].

Aim

The purpose of this paper is to present the current state of the nation-wide PAEHR implementation in Sweden, to describe the challenges in the implementation, and finally to discuss the need for patient-centered evaluation of PAEHRs as proposed in the PACESS project.

Methods

The results presented in this paper are based on a retrospective analysis of national projects in Sweden focusing on the design, implementation and evaluation of PAEHR. The authors have been engaged in the research projects described below over the past 7 years, and have had opportunity to observe the debate and decision making processes both in national and regional organizations. We base our description of the implementation projects and e-services on our participation in the projects as well as on publicly available resources provided by e.g. Inera regarding the current usage of the national e-services [13]. Over the years, the authors have also organized several international workshops, where challenges of implementing these types of eHealth services have been

discussed [14], [15], leading up to the analysis presented in this paper. The authors are part of the DOME consortium from which researchers of six Swedish universities to date have published over 50 articles on the introduction of the PAEHR (<https://domeprojekt.wordpress.com/>).

SUSTAINS

Already in 1997 Uppsala County Council (UCC) in Sweden started a project with the aim to give patients access to their medical data. The project was called Sustains and had financial support from the European Commission [16]. In 2002 a pilot study of PAEHR was run in Uppsala County, however it ran into legal problems and was shut down by the Data Inspection Board as the legislation at the time did not allow individuals to have online direct access to their medical records [17]. However, ensuing discussions resulted in the introduction of the Patient Data Act in 2008, which solved this problem. Work continued, and as part of the EU-project Sustains [17] (ICT-PSP 297206, during 2012-2014), UCC extended the deployment of public eHealth services, and all 300 000 patients were given access to the PAEHR in 2012.

The PAEHR Journalen

'Journalen' was originally developed and deployed in Uppsala County [17]. From 2015, the national eHealth organization (Inera AB) has the responsibility for development and maintenance of all public eHealth services and the national patient portal, further described in the Results section. Through this portal 'Journalen' is currently (December 2016) accessible by the citizens of 17/21 county councils.

My Care Pathways

The Swedish research project "My Care Pathways" [18] was another project running in parallel with the EU Sustains project also focusing on giving patients access to information from their medical records. This project aimed to create new mobile public eHealth services that allow patients to follow, own, and manage their care process-related information. The project also aimed to adapt and further develop the National Health Information Exchange (HIE) platform for citizen e-services and provide an open software development kit (SDK) for developing new e-services [19]. Several patient groups were involved in a patient-centered design process; stroke patients [20], lung cancer patients [21], and hip surgery patients [22].

Vårdhändelser

In parallel with the Sustains project, the My Care Pathways project also developed e-services for patients, the most notable called 'Vårdhändelser'. This service included similar health information as 'Journalen', but presented it to the user in the form of a time-line that was intended to be both backward and forward looking, indicating planned events, and including information from the patients' own health devices.

Results

In this section, a brief overview of the Swedish implementation of the PAEHR is presented, followed by challenges identified as essential during the implementation process. Finally, we introduce the PACESS project, which proposes that a patient-centered evaluation approach of PAEHRs is necessary to complement research in this area.

One way to access healthcare for the citizens – 1177.se

Although the county councils are autonomous and could prioritize which eHealth services to focus on, there is a national decision that patients should only have one way in to healthcare [4]. Thus, a patient portal '1177.se' is available for everyone seeking healthcare or health-related information in Sweden. The patient portal consists of three parts;

1. *1177 on the phone* - a telephone advice service reached through the national phone number 1177,
2. *1177.se on the web* - without authentication the public can access and search among information about illnesses, symptoms and treatments, as well as information about healthcare in the region. The virtual portal is national, but each region or county council in Sweden can adapt the information to its inhabitants.. As a user, it is also possible to choose which region you would like to view and change between regions..
3. *1177.se personal e-services* – after authentication (using a nationally approved BankID) you have access to personalized e-services where you can e.g. add your primary care centres or hospital units, and send secure messages to them. Depending on what e-services the care providers of your region offer, you may also e.g. request, reschedule or cancel appointments, renew prescriptions and access documents such as sick-leave.

Sweden has approximately 10 million inhabitants whereof 34.1 % (N=3 354 806) have created their own account for the 1177.se portal to use the personal e-services [13]. In September 2016, 1 427 576 log-ins were made to the personal e-services, and as an example 26 943 appointments were booked online [13].

Figure 1 shows an example of the different eHealth services a user can access via 1177.se (in this case the first author's own view when logged in). It is also possible to act as a proxy for children under 13 years, so parents can manage their children's appointments and contacts with healthcare.

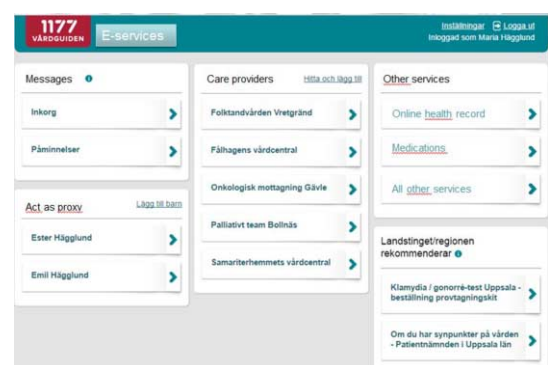


Figure 1– An example of a citizen view of the patient portal 1177.se after valid authentication (partially translated).

PAEHR in Sweden – retrospect and current challenges

Sweden has a decentralized healthcare system allowing for regional decisions.

For some time, both Journalen and Vårdhändelser (the alternative PAEHR service developed within My Care Pathways) were accessible to patients through 1177.se.

However, connecting to the respective e-service as an information provider was a costly process, and most care providers chose to publish their information to *either* Journalen or Vårdhändelser. This caused a fragmented view for patients; if you visited care providers that used the different services, some of your information would be visible through Journalen, and some through Vårdhändelser. Finally, all county councils agreed that Journalen is the national PAEHR service, to create one common view for the patients. Journalen has also been migrated from its UCC platform to the National HIE platform which also increased the incentives for the other, self-governing, care providers to connect their EHR systems.

Today (December 2016), 17 of the 21 Swedish counties have given access to EHR data through Journalen, and the service is expected to be truly national in the near future. Figure 2 gives an overview of the Swedish counties that have currently implemented and provide access to this service (in blue).



Figure 2—Implementation of PAEHR in Sweden, Dec. 2016.

The maturity of the implementation varies among the counties, e.g. Stockholm County Council, one of regions with most inhabitants in Sweden, have begun pilot testing during late 2016 and will implement the service throughout the county during 2017, whereas UCC has had Journalen in place for more than 5 years.

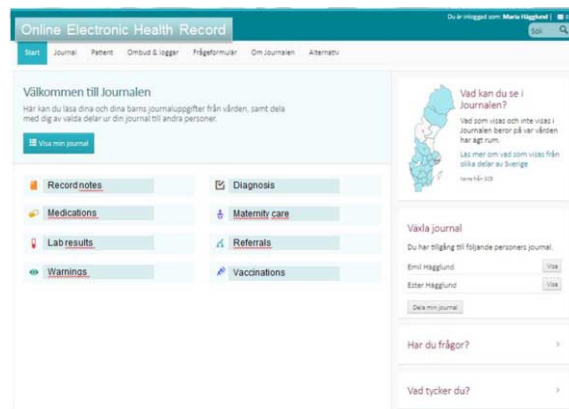


Figure 3—Example of PAEHR view (partially translated – Journalen is currently only available in Swedish).

Currently, when a patient is logged into 1177.se, and chooses to access the PAEHR the user finds the service shown in Figure 3. Journalen can contain notes from the EHRs (from all healthcare professions and all regions), a list of prescribed medications, lab results, warnings, diagnosis, maternity care records, referrals and vaccinations. In addition, there is a log list showing everyone that has accessed the record. The patient also has the possibility to share their EHR with anyone they choose, e.g. a close relative or an agent, and parents can access their children's records until the age of 13.

Agreeing on a national regulatory framework

One of the major challenges in the implementation process has been to achieve national consistency in the guiding rules for PAEHR. A National Regulatory Framework (NRF) was developed to support the deployment of the PAEHR in the 21 regions. However, from the point of view of the patient, the first version of the NRF was less successful as it contained electable paragraphs that were interpreted and applied differently in the counties [23], [24].

The PAEHR is one of the most important services for the patients and consequently the NRF needed an update [23]. In the recently launched second version of the NRF, the goals of the European and national eHealth strategies are used to create a number of principles, thereby giving the citizens the same opportunities regardless of where you live, where and when you seek care. Further, it should be possible for the regions to update their solutions to adhere to version 2.0 by 2020 [24].

Fragmented views of healthcare data

Despite the national HIE platform and the intention to provide patients with a complete overview of their health related data, the view remains fragmented depending on where, when and why a patient seeks treatment [23], [24]. There are important differences in how much information each care provider gives access to. Figure 4 provides an example of what information types some of the counties have chosen to provide (not including all information types and counties). For a complete list, please visit Inera's web site [25].

	Blekinge	Dalarna	Jönköping	Kalmar	Kronoberg	Norrbottn	Skåne	Södermanland	Uppsala
Notes									
Diagnosis									
Log reports									
Medications									
Lab results									
Referrals									
Warnings									
Vaccinations									

Figure 4—Examples of information provision from 9 of the 17 Swedish counties connected to Journalen, Dec 2016.

Local initiatives

Despite the national eHealth strategy, development and implementation of PAEHR in Sweden is based on local initiatives. Research and development projects, although funded by European and national funds, have been run in different regions in Sweden and decisions on a national level have only been made when absolutely necessary. Whether a top-down or a bottom-up approach to national eHealth implementation is best can be debated, but we will here describe some challenges observed in the implementation.

Different PAEHR services; different local initiatives running in parallel resulted in different solutions for accessing your health-related data. This would not have been a problem if all healthcare providers made their data accessible through both e-services, but due to cost and maintenance issues this was not feasible.

Poor coverage; although the national HIE platform aims to make all health-related data accessible, there are still many health information systems that are not connected to the platform. This can be due to technical issues, but also costs related to the connection can play a part – especially for smaller private care providers. Over time, hopefully this issue will become less of a problem as more and more care providers connect to the national platform.

Inequitable information provision; the different interpretations of the NRF causes uneven information access depending on which care providers you have visited. A patient that moves between care providers and/or counties, which is quite common, risk misunderstandings or frustration as information from parts of their care process is completely missing (when care providers are not connected) or partly missing (when care providers give limited access).

Overview of information; it is difficult for the user to get an overview of what content they can expect to be accessible from their care providers. All functionality is available, regardless of whether there is any information available or not, and if you have your lab results from one care provider but not from another it might be confusing. This causes many support errands from users asking where their information is [24].

Resistance from healthcare professionals

Patients' online access to their health records is a much debated topic. Although the research done regarding these types of services indicates that patients' experience mainly benefits [11], the fears among healthcare professionals remain high [5]. Hypotheses as to why this is so intimidating are many, but a prominent one stresses the power balance between patients and healthcare professionals as an explanation to the reluctance to share information with patients. Another angle is the paternalistic structures of traditional healthcare. Empowerment has been acknowledged as an alternative to compliance in order to guide the provider–patient relationship [26]. In the more traditional compliance-oriented approach to healthcare, patients are seen as the recipients of medical decisions and prescriptions, whereas the empowerment-oriented approach views patients as being responsible for their actions and their consequences. These perspectives clashed in a recent debate regarding a patient finding out about her cancer through the PAEHR [27]. Many voices were raised to condemn the PAEHR, arguing that it put patients at risk and that certain information needs to be blocked to protect patients. Others defended the right to decide for oneself if and when to read the PAEHR, and claimed that patients can handle these choices.

Long-term assessments are not prioritized

Evaluations of the effects of PAEHR have so far been project based and there's a lack of long-term evaluation planned. Within the DOME-consortium, researchers from different disciplines collaborate to research the impact of PAEHR, but financial support is scarce. Project-based assessments do not cover long-term effects and there are many questions that remain to be answered. Is the PAEHR a means to empower all patients or is it a tool appropriate for those who are already empowered? How can we address the needs of different patients, with different levels of health and digital literacy to navigate the online healthcare world? And what is the role of healthcare professionals in this process?

Limited support for patient participation and communication

One of the expected benefits of PAEHR is to increase patient participation in healthcare. Having access to one's own data is an important first step, but participation requires more. A

dialogue and collaboration between patients and healthcare professionals is required, and PAEHR does not automatically create this – in Sweden, the implementation has rather been performed under the premises that this is a tool for patients, so healthcare professionals will not need to change their way of working.

Focus in the PAEHR is also very much on giving patients access to information, not to support a two-way information exchange. Basic forms have been implemented in Uppsala that enables patients to fill out questionnaires that healthcare ask for, and a functionality for patients to comment on notes in the health record is also available – but not implemented or used outside Uppsala. In order to support patient participation and communication more interactive ways to exchange information would be required. In addition, integration of data from personal health apps or self-trackers could also be useful.

Discussion

Sweden has a decentralized healthcare system and the county councils and regions have great autonomy. Despite the national eHealth strategy, development and implementation of PAEHR in Sweden is based on local initiatives. Research and development projects, although funded by European and national funds, have been run in different regions in Sweden and decisions on a national level have only been made when absolutely necessary. Whether a top-down or a bottom-up approach to national eHealth implementation is best can be debated, but we have described some challenges that have been observed in the Swedish implementation; agreeing on a national regulatory framework, fragmented views of healthcare data, resistance from healthcare professionals, limited support for patient participation, and lack of longterm evaluations.

Healthcare professionals' resistance to PAEHR is also well-known from international literature [10]. When PAEHR is implemented, actual experiences are however mainly positive, especially patients experiences. In Sweden, despite the decentralized healthcare, decisions to implement PAEHR have mainly been politically driven, and resistance from healthcare professionals have often been strong. Health and digital literacy are often listed as risks, as are concerns that information in the PAEHR will cause worry for the patient. However, we argue that patients need to be allowed to make the decision of whether or not they want to use the PAEHR themselves, that actual patient experiences is an essential part of evaluating PAEHR's.

In order to harness the powers of eHealth, we need a better understanding of how roles, relationships and organizational structures are affected on micro, meso, and macro levels through the introduction of such eHealth services. In the recently funded research project **PACCESS**, the main objective is therefore to explore the impact of information access and innovative eHealth on patients and healthcare. 2 work packages are planned; (1) long-term evaluation of PAEHR in Sweden, and (2) co-creation of innovative e-services that empower citizens. Throughout the project we will establish a theoretical framework to explain how implementation of eHealth services impact different patient groups and healthcare on different levels of the health system. Furthermore, patient-centered quality indicators will be created to facilitate joint assessment of deployment and use of PAEHRs.

Conclusion

The Swedish path to a nationally available PAEHR has been long and challenging, and still we're not quite there yet. Challenges that remain have to do with local differences in the implementation that lead to fragmentation and unequal access to information. Initiatives have been taken to reconcile some of the problems, e.g. an updated national regulatory framework for PAEHR, but further efforts are needed to evaluate the implementation from a patient perspective.

Acknowledgements

The work presented in this paper has received funding from several sources; VINNOVA – Swedish Governmental Agency for Innovation Systems through “My Care Pathways” (2011-02536), and “DOME - Deployment of Online Medical Records and eHealth Services” (2012-02233). SUSTAINS was supported by the European Commission (No 297206) and FORTE – the Swedish Research Council for Health, Working Life and Welfare supports “PACESS” (2016-00623). We would also like to thank Inera AB (www.inera.se) for providing the data on the usage of the national e-services.

References

- [1] S. Koch, T. Hasvold, A. W. Kushniruk, A. Marcelo, and G. Kouematchoua Tchuitecheu, “eHealth - nurturing patient empowerment? State of the art and reflections from four continents. In: Panel discussion. Medinfo 2010, Cape Town, South-Africa.
- [2] S. Koch, “Improving quality of life through eHealth-the patient perspective,” *Stud. Health Technol. Inform.*, vol. 180, pp. 25–29, 2012.
- [3] F. Karlsson, J. Holgersson, E. Söderström, and K. Hedström, “Exploring user participation approaches in public e-service development,” *Gov. Inf. Q.*, vol. 29, no. 2, pp. 158–168, 2012.
- [4] Ministry of Health & Social Affairs., “National eHealth – the Strategy for Accessible and Secure Information in Health and Social Care,” 2010.
- [5] I. Scandurra, A. Jansson, M.-L. Forsberg-Fransson, and T. Ålander, “Is ‘Patient’s Online Access to Health Records’ a Good Reform? – Opinions from Swedish Healthcare Professionals Differ,” *Procedia Comput. Sci.*, vol. 64, pp. 964–968, 2015.
- [6] M. Hägglund and S. Koch, “Commentary: Sweden rolls out online access to medical records and is developing new e-health services to enable people to manage their care,” *BMJ*, vol. 350, p. h359, 2015.
- [7] K. Mandl, P. Szolovits, and I. Kohane, “Public standards and patients’ control: how to keep electronic medical records accessible but private,” *BMJ*, vol. 322, pp. 283–7, 2001.
- [8] D. Wiljer, S. Urowitz, and E. Apatu, “Canadian committee for patient accessible health records. Patient accessible EHR: exploring recommendations for successful implementation strategies,” *J Med Internet Res*, vol. 10, 2008.
- [9] M. Tiik, “Rules and access rights of the Estonian integrated e-Health system,” *Stud. Health Technol. Inform.*, vol. 156, pp. 245–256, 2010.
- [10] S. De Lusignan, F. Mold, A. Sheikh, A. Majeed, J. C. Wyatt, T. Quinn, M. Cavill, C. Franco, U. Chauhan, H. Blakey, N. Kataria, T. N. Arvanitis, and B. Ellis, “Patients’ online access to their electronic health records and linked online services: a systematic review in primary care,” *BMC open*, vol. 4, no. e006021, 2014.
- [11] C. Grunloh, A. Cajander, and G. Myreteg, “‘The Record is Our Work Tool!’-Physicians’ Framing of a Patient Portal in Sweden,” *J. Med. Internet Res.*, vol. 18, no. 6, p. e167, Jun. 2016.
- [12] S. R. Jilka, R. Callahan, N. Sevdalis, E. K. Mayer, and A. Darzi, “‘Nothing About Me Without Me’: An Interpretative Review of Patient Accessible Electronic Health Records,” *J. Med. Internet Res.*, vol. 17, no. 6, p. e161, Jun. 2015.
- [13] “Inera Invånartjänster,” 2017. [Online]. Available: <http://www.inera.se/INVANARTJANSTER/>.
- [14] M. Hägglund, O. Heinze, M. Pobiruchin, I. Scandurra, and M. Wiesner, “Report on the HEC 2016 Workshop : Opportunities and Challenges of Consumer-centric eHealth Services – An Interdisciplinary Workshop,” 2016.
- [15] I. Scandurra and M. Hägglund, “Developing and implementing national eHealth services for patients – an interactive exploration of challenges and potential solutions,” 2015.
- [16] B. Eklund and I. Joustra-Enquist, “Sustains - direct access for the patient to the medical record over the Internet,” *Stud Health Technol Inf.*, vol. 100, pp. 182–189, 2004.
- [17] L. Lyttkens, “Support Users To Access Information and Services - Final Pilot Evaluation Uppsala County Council,” 2014.
- [18] N. Lundberg, S. Koch, M. Hägglund, P. Bolin, N. Davoody, J. Eltes, O. Jarlman, A. Perlich, V. Vimarlund, and C. Winsnes, “My Care Pathways - creating open innovation in healthcare,” *Stud Heal. Technol Inf.*, vol. 192, pp. 687–91, 2013.
- [19] “Health Innovation Platform (HIP).” Available: <http://www.healthinnovationplatform.se/en/>.
- [20] M. Hägglund, N. Davoody, P. Bolin, C. Winsnes, N. Lundberg, V. Vimarlund, and S. Koch, “Stroke Patients’ Needs for Access to Care Process Related Information,” in *Medicine 2.0*, 2012.
- [21] M. Hägglund, P. Bolin, and S. Koch, “Living with Lung Cancer – Patients’ Experiences as Input to eHealth Service Design,” *Stud. Health Technol. Inform.*, vol. 216, pp. 391–395, 2015.
- [22] M. Hägglund, P. Bolin, and S. Koch, “Experiences as input to eHealth design - a hip surgery patient journey case,” *Stud. Health Technol. Inform.*, vol. 210, pp. 672–674, 2015.
- [23] I. Scandurra, L. Lyttkens, and B. Eklund, “Implications of Swedish National Regulatory Framework of the Patient Accessible Electronic Health Record,” *Stud Health Technol Inf.*, vol. 228, pp. 695–699, 2016.
- [24] I. Scandurra, M. Pettersson, B. Eklund, and L. Lyttkens, “Analysis of the Updated Swedish Regulatory Framework of the Patient Accessible Electronic Health Record in Relation to Usage Experience,” *full Pap. Submitt. to Medinfo*, 2017.
- [25] Inera, “Vad visar vårdgivare?”. Available: <http://www.inera.se/TJANSTER--PROJEKT/Journalen/Vad-visar-vardgivare/>.
- [26] I. Aujoulat, W. d’Hoore, and A. Deccache, “Patient empowerment in theory and practice: polysemy or cacophony?,” *Patient Educ. Couns.*, vol. 66, no. 1, pp. 13–20, Apr. 2007.
- [27] M. Lindgren, “Hon fick cancerbesked via nätjournal – trots rådrum i Jönköping.” *Sjukhusläkaren*, 2016.

Address for correspondence

Maria Hägglund, Karolinska Institutet/LIME, Tomtebodavägen 18A, 17177 Stockholm, Sweden. Email: maria.haggglund@ki.se, Phone: +46 (0)70 718 041

What Do They Mean by “Health Informatics”? Health Informations Posts Compared to Program Standards

Josette F Jones, RN, PhD^a, Enming Zhang, MS^a, Anand Kulanthaivel, MIS^a, Shilpa Katta^a

^a Department of BioHealth Informatics, Indiana University, Indianapolis, Indiana, USA

Abstract

There is a lack of alignment between and within the competencies and skills required by health informatics (HI)-related jobs and those present in academic curriculum frameworks. This study uses computational topic modeling for gap analysis of career needs vs. curriculum objectives. The seven AMIA-CAHIIM-accepted core knowledge domains were used to categorize a corpus of HI-related job postings (N=475) from a major United States-based job posting website. Computational modeling-generated topics were created and then compared and matched to the seven core knowledge domains. The HI-defining core domain, representing the intersection of health, technology and social/behavioral sciences matched only 45.9% of job posting content. Therefore, the authors suggest that bidirectional communication between academia and industry is needed in order to better align educational objectives to the demands of the job market.

Keywords:

Employment (D004651), Curriculum (D003479), Medical Informatics (D008490)

Introduction

The field of health informatics (HI) is a relatively new area of study and practice closely related to, and sometimes used interchangeably with *clinical informatics*, *medical informatics*, and *biomedical informatics*. Currently only the American Medical Informatics Association (AMIA) and the Commission on Accreditation for Health Informatics and Information Management Education (CAHIIM) provide a curriculum competency framework [1] describing the minimum knowledge and skills graduates of a HI program must attain. No governmental entity provides an official labor classification for professional in health informatics nor is there an agreement on the job titles.

The question arises if the competencies set forth by the curriculum framework are in harmony with the knowledge and skills required by the job market; Kulikowski et al [2] brought in one of the broadest definitions of health informatics then coined as *biomedical informatics (BMI)*. In this 2012 white paper, the authors proposed that core competencies in BMI would include basic biomedical science, information technology, computing, professionalism, and knowledge of the social science of information usage. More recently, Fridsma [3] published a viewpoint in the *Journal of the American Medical*

Informatics Association (JAMIA), using the concept of HI and no longer BMI to refer to the domain. HI is described as “intersections creating a continuum” ([3], p. 855) between various knowledge domains of application including healthcare, biosciences, computing, and social sciences and ranging from a disease management to public health applications and research.

The broadly stated competencies described in the 2012 AMIA white paper [2] were not yet in a form usable for formal accreditation processes. Since 2016, the AMIA Accreditation Committee (AAC) has been reframing and redefining through an iterative process the curriculum requirements as graduate outcomes. The resulting revisions [1] set forth 10 foundational domains, each with accompanying knowledge, skills, and attitudes necessary to succeed as health informatics professionals in an ever-changing job market. The three base domains are: Health (“F1”; biomedical and health sciences), Information Science & Technology (“F2”; methods and technologies for storage and exchange of information), and Social & Behavioral Science (“F3”; psychology, sociology, organizational behavior).

Three second-level intersected domains are formed by the joining of any two of the above base domains: F1 and F2 combine to form Health Information Science & Technology (F4), F2 and F3 combine to form Human Factors & Sociotechnical Systems (F5), and F1 and F3 combine to form the Social & Behavioral Aspects of Health (F6).

The unique core of health informatics is at the intersection of F1, F2, and F3; the intersection is known as “Social, Behavioral, and Information Science and Technology Applied to Health” (F7), and its specific knowledge, skills and attitudes are considered the defining and differentiating foundation HI program. [1]

The AAC added three domains: Professionalism, leadership, and interprofessional collaborative practice. The relationships and intersections among the original seven domains are visualized in Figure 1. The AAC’s three added domains, although important, are not considered in this analysis as these domains are considered intrinsically required by most jobs, regardless of field. The foundational domains of HI education are depicted in figure 1.

Yet still the question remains: do the knowledge, skills and attitudes set forth in the most recent recommendations [1] cover the demands of the job market?

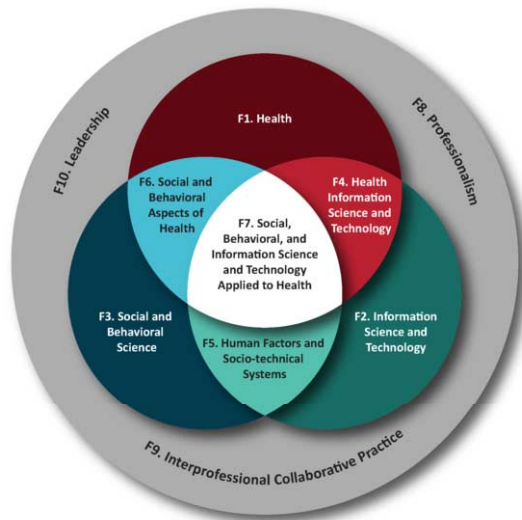


Figure 1. Venn Diagram of the Competency Domains Proposed by the AAC, 2016 [1].

Objective

Due to the inconsistencies in job descriptions and the recent revisions in HI program outcomes and related competencies, the need arises to perform a gap analysis between the knowledge and skill domains identified for the HI curriculum and actual posted job requirements.

Methods

NLP: Computational Topic Modeling

A convenience sample of US-based job postings made as of mid-November 2016 from the website Indeed.com [4] was searched and mined using the Publisher Toolkit's application programming interface [5] connected to a proprietary Python-based script. To extract the HI-related postings, the terms “health informatics”, “clinical informatics”, and “medical informatics” were used to filter the resulting corpus. The resulting filtered data were then organizing into individual documents, each document containing one job posting.

MALLET (Machine Language Learning Toolkit) [6] is an open-source topic modeling software tool that utilizes an algorithm that computes baskets-of-words (named topics) that frequently occur across a set of documents. MALLET assigns individual weights to each basket-of-words within each document. For this study, MALLET output was exhausted at 30 topics co-occurring across these 475 job-posting documents.

Qualitative Second-Level Matching

Following MALLET categorization, *second level matching* was performed: Two HI graduate students independently assigned one of the seven core AMIA/CAHIIM knowledge domains [1, 2] to each MALLET-generated keyword basket. The inter-rater agreement was calculated at this stage (23/30 topics; 76.7%). Disagreement was resolved by mutual adjudication. A similar process was used for labeling spurious topics such as those referring to marketing publicity.

The topic weights were then summed to determine the weight of each individual domain from F1 through F7, including the spurious topics, within each job posting. For example, if a job matched to the three MALLET generated topics for foundational domain F2 at strengths 0.0528 (software development), 0.000140 (data science), and 0.00044 (software application) respectively, the overall weight for domain F2 would be 0.0534.

Binary Model Generation

The matrix was simplified into a binary matrix – potential combinations of the foundational domains excluding the nucleus of HI – by only including relative job-domain strengths that exceeded 0.242 (equivalent to one standard deviation above the median). These strengths were coded as “1”, while lower associations were coded as “0”.

Reclaiming Missed Postings

In order to verify that all potential HI job postings were extracted from the original list, we analyzed those jobs that did not match F7 in the binary model but matched entirely one of the combinations below:

- F1 + F2 + F3
- F3 + F4
- F1 + F5
- F2 + F6
- F5 + F6
- F4 + F5
- F6 + F7

Results

NLP: Computational Topic Modeling

The retrieval procedure (downloading from Indeed.com) yielded 475 job postings that matched the search query. MALLET determined 30 topics. In order to determine the relative weight of each domain across the corpus, the document-topic strength file generated by the software was analyzed.

An example of the data job-topic data generated by MALLET is seen in the following table:

Topic	T00	T01	T02
Job 0	0.0001	0.0624	0.0528
Job 1	0.0001	0.1594	0.0002
Job 2	0.0000	0.0004	0.0001

Table 1. Document (Job posting)-to-Topic Weights.

Qualitative Second-Level Matching

Each topic was qualitatively assigned a domain by the panel of two graduate assistants. A list of topics, along with the qualitatively assigned knowledge domains, is available in Appendix A.

During the rating of the MALLET-generated topics, the graduate assistants also excluded by consensus any topics they determined to be unrelated to actual core knowledge domains. Specifically, 8 out of 30 topics generated were deemed by consensus to be spurious as shown in the table below.

Reason deemed spurious	# of Topics
Exclusively advertised the employer	5
Exclusively described employee physical characteristics (physical labor capabilities) and not employee competencies	2
Consisted entirely of mandatory legal language (equal opportunity statement, etc.)	1

Table 2. Spurious Topics and Frequencies.

The topics were then collated into cumulative strengths that represented the strength of each domain across the postings. The median of all job-domain strengths was found to be 0.024; the standard deviation of job-domain strengths was 0.218. An example of cumulative strengths by domain for job postings 0, 1, 2, and 3 across F4, F5, and F6 is below.

Job	F4	F5	F6
0	0.000	0.199	0.594
1	0.000	0.027	0.227
2	0.000	0.001	0.997
3	0.001	0.002	0.989

Table 3. Job Posting Cumulative Strength by Domain, Examples.

Binary Model Generated

The full binary job-domain relationship table was generated; a truncated example of the results (derived from Table 3's domain strengths) is in the below table.

Job	F4	F5	F6
0	0	0	1
1	0	0	0
2	0	0	1
3	0	0	1

Table 4. Binary Representation of Job Posting Cumulative Strengths by Domain.

The binary model also revealed the following match rates for each domain:

Domain	Postings Matched	
	N	%
F1	39	8.2%
F2	44	9.3%
F3	12	2.5%
F4	29	6.1%
F5	43	9.1%
F6	376	79.2%
F7	40	8.4%
(Spurious Topics)	173	36.4%

Table 5. Binary Results for Matching of Job Postings to Topics Related to Each Competency Domain.

The domain (F7) considered the unique core of HI matched 8.4% postings at an overall strength of 0.242 or higher. 36.4% of postings matched spurious topics, although only 43 (8.8% of corpus total) job postings matched *only* the spurious topic set. The most job descriptions (79.2%) match the competencies expressed in foundational domain F6: social and behavioral aspects of health systems. One job posting (0.2%) did not match any of the topics at a binary threshold of 0.242, meaning that the total adjusted match rate for *any* non-spurious topic was 90.9% (N=433).

Reclaiming Missed Postings

The lack of postings (8.4%) that initially matched the F7 domain is of concern. This deficit, however, was amended by *reclaiming* postings that missed classification as F7. In order to increase the match rate for job postings that potentially related to the F7 domain, combinations of co-matches between topics that together would make up F7 were sought, increasing the F7 match rate from 8.4% (N=40) to 24.2% (N=115).

Domain Combination	N	% of corpus
F1 + F2 + F3	0	0.0%
F3 + F4	0	0.0%
F1 + F5	1	0.2 %
F2 + F6	33	7.0%
F5 + F6	23	4.8%
F4 + F5	0	0.0%
F6 + F7	18	3.8%
Recovered Total	75	15.8%
(F7 initial)	40	8.4%
Sum: Total possible matches	115	24.2%

Table 6. Job Postings Recovered by Combinations of Domains that Could Represent F7.

Discussion

The analysis revealed certain matches between the curriculum competencies and the skills and knowledge desired in 475 real life job postings on *health informatics* and/or related terms.

The least frequently matched domain was Social & Behavioral Sciences (F3), matching 2.5% of job postings. This finding implies that employers looking for HI-related professionals may seek fewer individuals who solely have social-behavioral science competency. The most popular competency domain correlated to these job postings, on the other hand, was observed to be Social & Behavioral Aspects of Health (F6), which matched 79.2% of all job postings in the binary model, indicating that combined competencies bridging Social & Behavioral Sciences with Health Sciences are in relatively high demand.

Match Rate for F7 – Social, Behavioral and Information Science and Technology Applied to Health

The reclaimed postings, while they only partially match to the core of HI [1,2], will still be considered related to Health Informatics.

The F7 match rate (maximum of 24.2% even after reclamation) remains lower than the F6 match rate (79.2%) suggesting that there exist many *health informatics*-related jobs that *employers describe as* requiring knowledge of social sciences and health while requiring little to no technology competency.

Use of Qualitative (Second Level) Matching

It also follows that qualitative *matching* analysis to interpret the *keyword baskets* generated by NLP software such as MALLET may have its pitfalls due to potential bias on behalf of the raters. Nonetheless, qualitative secondary matching has been shown of use in health related applications including coding medical concepts from clinical free text entries [7] and classification-annotation of mentions of pharmaceutical treatments [8] the authors of both of these studies implore that qualitative analysis is required following computational NLP.

Limitations: MALLET Analysis

It is also noted that there is required a future qualitative analysis in terms of quality assurance of the MALLET-generated results. In the future, other analyses should be undertaken in order to study this and similar corpora: Further study may include synonym-based (synset) literal term searching, as well as qualitative analysis of a reduced corpus in order to perform an expert categorization of job postings.

Limitations: Spurious Topics & Posting Bias

The presence of what the authors term *spurious topics* (those that exclusively advertised the employer, location, or legal requirements) must also be addressed, as these spurious topics matched to 36.4% of all job postings in the binary model. In fact, 8.8% of all job postings matched *only* the spurious topics.

The corpus that was searched and processed may therefore represent a biased sample. Specifically, it likely represents the wording chosen by the writers of the job postings. While the postings did contain information about the core knowledge domains required to satisfy job demands, it is also observed via the strength and content of spurious topics that significant amounts of advertising and legal information was present in the job postings. Furthermore, if there is any inaccuracy or inexperience on behalf of those writing the job postings, such error cannot be taken into account by the study design at hand.

Limitation: Incomplete Coverage of Domains

The authors also note that a limitation exists in the coverage of all competency domains and their expressions of knowledge, skills and attitudes.. While the CAHIIM-AMIA competency classification includes 10 domains, only the 7 domains from the overlapping Venn diagram were covered in this proof of concept study. Future studies will need to cover all 10 competency domains in order to better match job requirements to curriculum competencies.

Conclusions

The study reported was able to create a connection between a major upcoming curriculum competency domains and real life job postings for health informatics and related areas. Computational topic modeling/NLP followed by qualitative (*second level*) consensus analysis has here shown the ability to process a large corpora (N = 475) of job postings and assign academically-meaningful topics to 433 (90.9%) of them.

Therefore, the authors recommend that NLP via computational topic modeling, followed by qualitative second-level analysis, be used to analyze further corpora of job postings in order to match them to curriculum frameworks. Such matching analysis is likely to reveal the similarities and differences between curricular needs and the job market; this information can be used to tailor curriculum frameworks as well as employer job postings.

Acknowledgements

The authors would like to acknowledge the AMIA Accreditation Committee for their revisions of the foundational domains of Health Informatics.

References

- [1] Commission on Accreditation for Health Informatics and Information Management Education (CAHIIM) & American Medical Informatics Association (AMIA). AMIA Accreditation Committee : Draft For Public Comment, 2016.
- [2] Kulikowski C, Shortliffe E, Currie L, Ekin P, et al. AMIA board white paper : Definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc* (2012), 931-938.
- [3] Fridsma DB. The scope of health informatics and the Advanced Health Informatics Certification, *J Am Med Inform Assoc* (2016), 855-856.
- [4] Indeed.com, Inc. Job Search. (Website), n.d.-2016. Retrieved from <http://www.indeed.com/>
- [5] Indeed.com, Inc. Publisher API. (Online Software Tool), 2016. Retrieved from <http://www.indeed.com/publisher>
- [6] McCallum AK et al. MALLET : A Machine Learning for Language Toolkit. (Software Tool). Amherst, MA : University of Massachusetts, 2002. Retrieved from <http://mallet.cs.umass.edu/>
- [7] Lin CH, Wu NY, Lai WS, Liou DM. Comparison of a semi-automatic annotation tool and a natural language processing application for the generation of clinical statement entries. *J Am Med Inform Assoc* 22 (2015), 132-142.
- [8] Merabti T, Abdoune H, Letord C, Sakji S, Joubert M, Darmoni S. Mapping the ATC classification to UMLS metathesaurus: Some pragmatic applications, *Stud Health Technol Inform* (2011), 206-213.

Address for correspondence

Corresponding Author: Dr. Josette Jones.

Email: jofjones@iupui.edu.

Mailing Address: 719 Indiana Avenue, Walker Plaza #319, Indianapolis, Indiana USA 46202.

Appendix A: Word Baskets & Topics with qualitative interpretation

Topic No.	Topic/Word Basket	Interpretation
T00	nursing colorado library school pharmacy city campus kansas hays researchers board state anschutz denver regional schools resources hit access umkc	Spurious
T01	health research informatics public sciences programs school position care medicine policy services education center program demonstrated population social staff medical	F6
T02	experience development software web team strong years programming java informatics technologies boston company equivalent developer applications gns cloud performance science	F2
T03	skills work experience ability knowledge information health management communication working team position informatics related required degree project provide projects including	F6
T04	business sales customer account market marketing develop product opportunities products strategies healthcare create drive strategy experience industry accounts key customers	F6
T05	research center data science bioinformatics computing computational biomedical scientific medicine school cancer biology texas collaborative expected genomics candidate postdoctoral university	F4
T06	management health information development technology support planning leadership administration including policies business operations experience initiatives program strategic objectives plans programs	F5
T07	infrastructure including pharmacy alliance tobacco application architecture results benefits develop network medicine university compute cancer monitor pci build technologies control	F5
T08	usf research position clinical florida universities tampa public opportunity working cover equal employees institute south system apply resume click top	Spurious
T09	develops maintains related management performs quality requirements ensures program works participates functions utilization serves reports procedures leads assists demonstrates services	F3
T10	status employment opportunity equal gender national disability protected sexual race color religion age origin veteran orientation identity applicants employer law	Spurious
T11	college georgia education students student community campus faculty school degrees university academic technology duties nursing coastal advising studies experience ccga	Spurious
T12	health information clinical informatics care data technology nursing science practice systems develop design applications administration management implement assist staff computer	F5
T13	data public health disease healthcare american surveillance heart association epidemiology prevention opportunity resume division fellowship duties city stroke impact control	F1
T14	philips health job sales company act care opportunity solutions home contact technology consumer application process applicants title employer clinical protected	Spurious
T15	university faculty teaching department information position program students graduate candidates science research qualifications applications professor informatics application professional college include	Spurious
T16	project manager united healthcare management experience specific ability electronic account skills implementation members role primary patients job diagnosis knowledge team	F6
T17	data analysis health experience business statistical reporting analytic analyst tools complex required research reports preferred years analytics requirements including database	F7
T18	healthcare business team solutions data clients work services teams project product organization client stakeholders lead analytics care ensure processes understanding	F6
T19	quality clinical health measures experience federal measure measurement programs review working project care research ngf cms state performance public agencies	F6
T20	healthcare sales digital quality customer clinical solutions business product leadership including customers strong products global company imaging willingness ambulatory technical	F6
T21	quality care health data improvement reporting healthcare performance activities provider improve meaningful providers reports medicare patient measures analysis national analytics	F6
T22	coding health information documentation icd required management knowledge registered program rhia ccs certification rhit education accurate medical director reimbursement codes	F6
T23	medical group payer building rules youâ€™ll ideas improve finance growing performance make review join relationship industry unitedhealth payers culture insurance	Spurious
T24	ability essential physical job perform employee functions demands pharmacy time required vision occasionally individuals including reasonable disabilities pounds telephone made	Spurious
T25	patient service patients department ensure field position safety policies medical reports staff daily procedures equivalent mount supervisor competencies customer current	F6
T26	data learning required machine analytics knowledge solutions mining predictive modeling language group decision iet models janssen processing statistics sciences scientist	F2
T27	clinical informatics system information systems workflow ehr user technology workflows michigan training end physician users specialist leadership patient data setting	F4
T28	systems software system technical application computer requirements support design development epic applications testing technology documentation solutions users services maintain user	F2
T29	clinical medical experience care preferred staff education quality patient practice required years ensure healthcare improvement duties physicians assigned minimum training	F1

Design of a User-Centered Voluntary Reporting System for Patient Safety Events

Hong Kang, Yang Gong

School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, USA

Abstract

As the third leading cause of death in the U.S., patient safety events (PSE) are difficult to control due to multiple inputs from healthcare providers, systems, or even patients. Inspired by the success of reporting systems in other fields, PSE reporting systems could be a good resource to share and to learn from previous cases. However, the success of such systems in healthcare is yet to be seen due to the low report quality and the lack of interoperability and communication. A knowledge-based and user-centered PSE reporting system is needed to organize the scattered knowledge and improve user-friendliness. We described the development of a knowledge base for patient falls, the most frequent PSE. Based on the knowledge base, user-centered design features were incorporated into the system to improve the reporting accuracy, completeness, and timeliness. This prototype holds promise in improving PSE reporting quality and facilitating human-computer communication.

Keywords:

Patient Safety; Medical Errors; Knowledge Base; Medical Informatics

Introduction

A patient safety event (PSE) is an event or circumstance that could have resulted, or did result, in unnecessary harm to a patient [1]. An estimated number of 15 million PSEs occurred in U.S. hospitals each year, as high as 40,000 per day, which exceeds the combined number from motor and air crashes, suicides, poisonings, and drownings [2]. With more than 251,000 annual deaths which are about 9.5% of all deaths, PSE, the third leading cause of death in the U.S. closely following heart disease and cancer, costs more than 9 billion dollars every year [3; 4]. Although most patients do not die from PSEs, they suffer from the PSEs for a long period of time or even the rest of their lives [5]. For example, patients may get a fracture after a serious fall in hospital or have a worsening medical condition by taking the incorrect dose of medications. PSEs, including near misses or close calls that are recognized before they actually occur, may be related to systems, operations, drug administration or any clinical aspect of patient care [6]. In the latest version of Common Definitions and Reporting Formats Version 2.0 (Common Formats, CFs) released by Agency for Healthcare Research and Quality (AHRQ), PSEs were categorized into nine subtypes including 1) blood or blood product, 2) device or medical/surgical supply including health information technology, 3) fall, 4) medication or other substance, 5) perinatal, 6) pressure ulcer, 7) surgery, 8) anesthesia, and 9) venous thromboembolism [7]. Different from diseases, which could be effectively controlled in accordance with clinical procedures, PSEs are difficult to control due to multiple inputs including healthcare providers, systems, or even

patients [8]. Therefore, PSE is a major threat to the healthcare quality.

Event reporting has been proven effective by many high risk industries such as aviation, nuclear, and rail industry, for improving safety and enhancing organizational learning from errors. In healthcare fields, PSE reporting systems would enable safety specialists to analyze events, identify underlying factors, and generate actionable knowledge to mitigate risks [9-11]. Dozens of PSE reporting system have been established based on this purpose. In the U.S., the Institute of Medicine (IOM) recommended using patient safety reporting systems (PSRS) [12] to evaluate why patients are harmed by health care [13]. AHRQ created the CFs [7] to help healthcare providers uniformly report PSE. Since 2000, at least 30 other PSE reporting systems have been established in the U.S., the initiatives to improve patient safety based on the common belief that data supports further learning and actionable knowledge. However, the success of such systems in healthcare is yet to be seen due to the identified barriers such as low report quality and the lack of interoperability and communication [14].

In addition to the no-blame no-shame safety culture that needs to be further enhanced, a lack of effective and efficient human-computer interaction (HCI) may largely account for the issues of low user acceptance and low-quality data currently confronting the PSE reporting systems [15]. Improved HCI in the systems may include individualized interfaces according to the user roles and requirements, increased sensitivity to the needs of the current clinical scenario, or even patient interfaces to enhance the patients' self-efficacy and awareness and thereby reduce PSEs. Our previous work indicated that the retrospective think-aloud user testing method is a useful usability evaluation method by which multidimensional measures can be synthesized to gain an insightful understanding of the usability in a voluntary patient safety reporting system [16]. However, generic PSE systems do not incorporate user-centered design (UCD), which is a major barrier to collect event data from frontline practitioners and to learn from previous events.

Thanks to the advancement of machine learning and web programming techniques, increasing UCD features have been incorporated into PSE reporting systems, such as spreadsheets, keyword searching, and automatic error correction. Nevertheless, investigating and learning from the reported events still largely rely on manual approaches due to the lack of an integrated view of PSE [17]. Developing a PSE knowledge base is necessary because it could acquire, organize and integrate PSE information and connect reported events to potential solutions. More importantly, the knowledge base would enable the users to view the system with an open and receptive mind. The story of *the blind men and the elephant* [18] is often told to PSE reporters. In the story, each man had touched only a part of the elephant thus making it impossible for him to know the whole animal. The same is true for

unintegrated reporting systems. The PSE knowledge base holds promise for organizing PSE knowledge and supporting advanced UCD features toward shared learning.

In our preliminary study [14], we developed a PSE similarity searching model by utilizing the semantic similarity measures on the PSE datasets of AHRQ WebM&M (Morbidity and Mortality Rounds on the Web) [19] and CFs [7]. Based on this model, we proposed a novel schema which can process the comparison tasks for PSEs and provide the reporters pertinent suggestions about solutions and prevention options for their cases. Patient falls, a subtype of PSEs, was chosen to assess the schema. As a follow-up study, this paper focused on the development of PSE knowledge base and UCD features. We developed a knowledge base for fall events, based on which we incorporated multiple UCD features into the reporting system.

Methods

Prototype a PSE knowledge base

The PSE knowledge base we prototyped was the collection of PSE reports, solutions, and the potential connections among them (Figure 1).

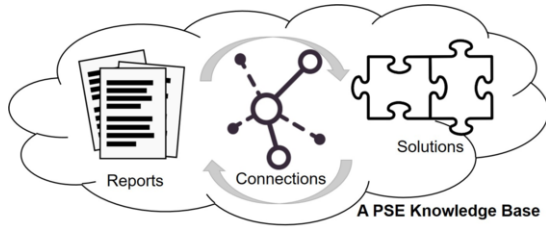


Figure 1—An infrastructure of PSE knowledge base

The knowledge base establishment was started from the most common PSE subtype, patient falls. In our preliminary study, we collected more than 7,000 fall reports from an institute from Patient Safety Organization (PSO) and developed a similarity searching model based on semantic similarity methods to make connections among reports, among solutions, and between reports and solutions [14]. To enrich the knowledge base for patient falls, we identified solutions for patient falls from multiple authoritative resources, such as the AHRQ WebM&M [19], Joint Commission Center for Transforming Health Care's Targeted Solutions Tool [20], Pennsylvania Patient Safety Authority [21], National Safety Council and National Patient Safety Agency's Patient Safety Observatory report [22], and synthesized them by building a connection between the entry-based solutions and the AHRQ CFs. These solutions were summarized and grouped into two types: general solutions (for all patient fall event reporters) and specific solutions (customized according to the reporting contents). We also initiated a survey with the Missouri Center for Patient Safety to evaluate and extend our solution entries. In the survey, we provided a text box following each solution for experts to comment on the adequacy of the solutions to the corresponding questions and answers in CFs. Five experts who are familiar with PSE reporting process and patient safety data participated in the survey. The contents of the solutions and their connections to the CFs were improved according to the expert comments.

Review current PSE reporting systems

To figure out the status quo of PSE reporting systems and propose an improved reporting system toward high-quality reporting, we investigated the current systems from peer reviewed publications and publicly accessible web pages.

Resources for publications included three databases: Ovid MEDLINE, PsycINFO, and Health and Psychological Instruments. Keywords including "patient safety event", "medical error/incident/event", "report/reporting system", "electronic report", "healthcare", "information system" were applied with different combinations to all field search (title, abstract, keywords, etc.). As this strategy may include articles with high sensitivity and low specificity, we set restrictions on the MeSH Subject Heading to match such term clusters as "Risk/Safety Management"/ "Quality of Health Care"/ "Quality Assurance" and "Patients or Medical Records Systems"/ "Computerized or Hospital Information Systems". We used Google search engine to identify publicly accessible PSE reporting systems and other information sources that contain substantial system descriptions (e.g., screen shots or demonstration videos) as supplemental information. Results from these additional materials were merged into the results of the literature review. Two domain experts filtered the initial results and generated a finalized review list.

Design a user-centered voluntary reporting system for PSE

The input quality of PSE reporting system relies on the UCD since UCD has been proven effective in user acceptance [16; 23]. Guided by the reporting quality-related factors derived from the review of current PSE reporting systems, we applied UCD to our reporting system with the support of the PSE knowledge base.

Results

A PSE knowledge base for patient falls

32 general and 137 specific solution entries for patient falls were determined after the survey results were analyzed. The survey also helped us assess the mapping rules between reports and solutions. Each specific solution entry will be displayed to the reporter only when the corresponding condition is met. The condition consists of combinations of users' answers to the 13 questions in the fall event CFs. Table 1 shows an example of the activation condition of a specific solution S_i .

Table 1—The activation condition of solution S_i

Solution S_i is activated when	
The h th answer of Question Q_{j1} is selected	
AND The k th answer of Question Q_{j2} is not selected	
$i=1,2,\dots,137; j=1,2,\dots,13$	

Three data quality dimensions for PSE reporting systems

48 unique PSE reporting systems in the United States, Netherlands, Canada, United Kingdom, Germany, Australia, China, and Japan were identified and reviewed by domain experts. Based on which, three data quality dimensions were defined as follows.

- **Accuracy:** the degree of proximity of a given PSE report to corresponding real world occurrences. The reporting accuracy is subject to user error and cognitive limitations in memory and reasoning, including but not limited to typographical errors, memory decay, casual attribution and hindsight biases. Accuracy of reporting could be improved if these contributing factors are incorporated into design consideration with good usability and functionality.
- **Completeness:** the degree to which a given PSE report includes necessary information describing the corresponding real world event so as to be

sufficiently valid for the purpose of analysis and generation of intervention. The completeness could be enhanced if its criteria are explicitly delineated and properly represented to the reporters with the help of interface features.

- **Timeliness:** the degree to which a PSE is reported in a timely manner for root cause analysis and generation of real time intervention. It can be enhanced by improving the efficiency of the reporting process and offering a smooth process to generate actionable knowledge as soon as the report is identified by reviewer.

A user-centered PSE reporting system

To improve the accuracy, completeness, and timeliness of PSE reporting systems, a set of UCD features determined in our preliminary studies [24; 25] were developed and incorporated into our system. Table 2 shows a summary of the available and under-development UCD features in our reporting system.

Table 2– UCD features of the proposed PSE reporting system toward high-quality reporting

UCD Features	Accuracy	Completeness	Timeliness
• Validator	✓	✓	✓
• Knowledge support and user feedback	✓	✓	
• User-friendly layout		✓	✓
• Role-based reporting and learning		✓	✓
• System interoperability	✓		✓
• Instant communication	✓		✓

(Features in an *italic font* are under development)

Validator

Multiple validators were incorporated into the reporting system. For example, the completeness validator can check whether all necessary fields have been filled by the reporter before the final submission to ensure the completeness of the report (Figure 2a). The spelling validator can identify spelling mistakes to avoid unnecessary accuracy loss during the similarity searching (Figure 2b). We initialized a terminology list in patient safety domain to standardize the words and phrases which may be applicable for reporters to choose from during reporting. The spelling validator was further improved toward text prediction based on the terminology list. Over time of individual and group use, nonstandard inputs are expected to be identified by this validator, and the possible standard terms will be prioritized for further selection (Figure 2b), which holds promise in improving the reporting accuracy, consistency and efficiency. Furthermore, the system can track new terms and update their frequencies to inform the system administrator periodically. The terms with high frequencies will be reviewed by domain experts and then put into an updated terminology list to enhance data entry quality.

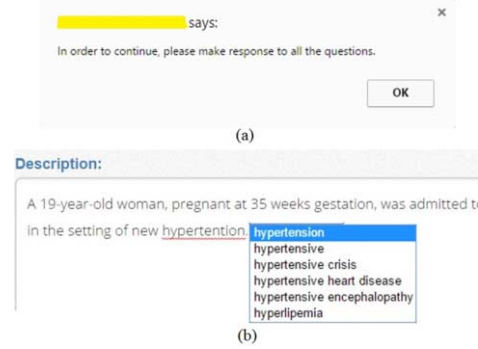


Figure 2– Screenshots for the validators. (a) Completeness validator (the IP address of local server was de-identified); (b) Validator for spelling and text prediction.

Knowledge support and user feedback

With the support of the PSE knowledge base, the user can either choose an existing case or report a new PSE as a query, and the system will retrieve similar cases and customized solutions for promoting shared learning.

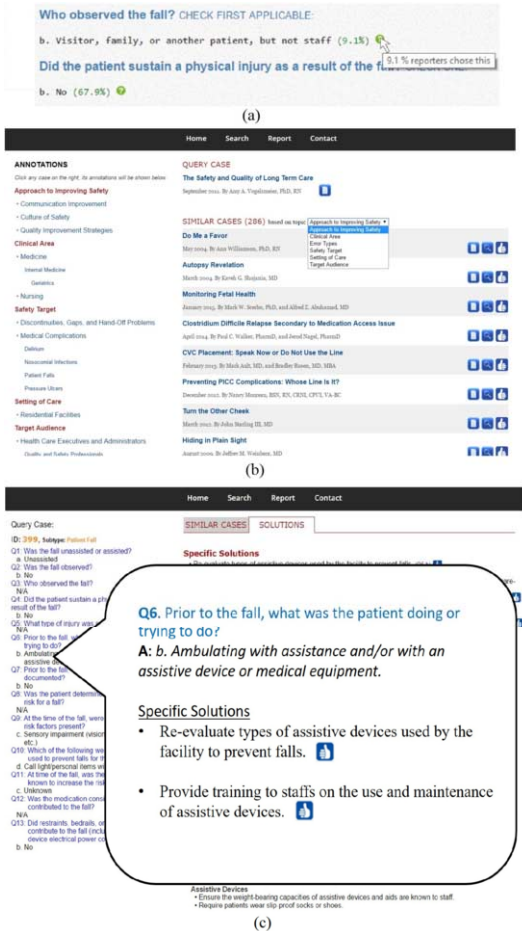


Figure 3– Screenshots for the knowledge support and user feedback. (a) Timely statistical analysis during reporting; (b) Providing similar cases for a new report; (c) Providing customized solutions based on user's reporting.

As shown in Figure 3, the knowledge support and user feedback are offered for fall events in our system. Reporters can receive timely notifications about the distribution of each answer option during reporting (Figure 3a), as well as similar cases (Figure 3b) and actionable solutions (Figure 3c) after reporting. The similar cases provide previous experiences and evidence-based suggestions from healthcare experts to prevent the potential consequences of the query event. The customized and actionable solutions can offer an interactive way to solve the current event. Users are allowed to send their feedback by clicking the thumb-up buttons to let the system know their preferences. The weights of the selected similar cases and solutions will be automatically enhanced in the similarity matrices. Therefore, the knowledge support is expected to become more precise enhanced by increasing number of user feedback.

User-friendly layout

The layout of user interface was an important factor we considered during the development. A user-friendly layout could improve the reporting system in terms of completeness and timeliness. For example, the hierarchical question layout during reporting reduces reporters' memory load for the particular task operation and decreases the likelihood of skipping correct answers.

Role-based reporting and learning

To provide customized solutions according to the various roles of healthcare providers (e.g., managers, clinicians, and staff) or even patients, we are classifying the solution entries into 3 categories: direct actions (for clinical staff to make specific actions), principles (unspecific actions, may be actionable for managers) and patients (patient actions). The system will retrieve similar cases and customized solutions based on the query and the reporter's role.

System interoperability

Lacking considerations on system interoperability and communication indicates a poor integration of event reporting procedure into clinicians' work flow, organizational quality control, and risk management process. Interfaces with other clinical applications (e.g., drug-drug interaction system) are under development.

Instant communication

Timeliness could be enhanced by instant communication between reporter and expert, and feedback access or notification. Both internal message module and live chat module will be developed to facilitate such communication.

Discussion

The evolution of PSE reporting systems

Similar to the evolution of electronic health record (EHR), PSE reporting systems started from an electronic copy of paper-based reporting forms. In this phase, the reporting systems can be viewed as a primitive alteration of paper-based reporting forms toward an intelligent reporting system. The use of drop-down lists, check boxes, or radio button replaces unnecessary free text boxes accelerates the electronic entry process and improves data accuracy by reducing data entry errors. The AHRQ CFs are designed to support operational systems at three levels: (1) support patient safety event reporting, which is currently a self-contained part of any EHR, (2) support surveillance based on the data derived from EHR and, (3) enhance analytics of safety and quality toward clinical decision support by linking PSE with EHR. Therefore, future PSE reporting systems should evolve toward knowledge-based and

user-centered systems which could improve reporting quality by offering timely knowledge support.

A shift from quantity to quality of event reports

When the quantity of reports is the only factor addressed in a safety culture discussion, an increase in event reports might be regarded as a reflection of an improved reporting culture, while others may consider a reduction in event reports as an indication of a safer environment. Nonetheless, underreporting, low quality and fragmented reports have not been adequately addressed in event reporting. We envision that the user-centered and knowledge-based design will revolutionize the traditional event report strategy, advancing from simply counting events into a new era of understanding, trending, integrating, and resolving the events through a synchronous and collaborative platform.

A no-blame no-shame culture beyond reporting system

Patient safety is as much about behavior, value and attitudes as it is about physical action. Another challenge for improving safety is to cultivate a no-blame no-shame culture. The characteristics of a positive safety culture include communication founded on mutual trust and openness, good information flow and processing, shared perceptions of the importance of safety, and recognition of the inevitability to error, etc. More efforts should be made to help healthcare providers understand their roles in improving patient safety. The first step may be to ensure patient safety is of high priority for each healthcare organization. Opportunities have to be created for people to freely state their opinions, and this openness then needs to be transferred to systems that allow all individuals to report and discuss. A no-blame no-shame culture will allow individuals to report and discuss in a comfortable atmosphere.

Limitations

All the assessments in this project were processed through expert review since there is no gold standard for PSE similarity measurement and solution recommendation strategy. Each expert might bring a different perspective which may result in bias toward the variation among similarity scores and solution mapping rules. For example, a physician may judge the similarity between PSEs by measuring severity, while a nurse may judge the similarity based on suggested solutions. The biases are inevitable but should have been minimized based on the common understanding of safety and quality. Thereby, we provided targeted introductions before every round of expert review and use face-to-face interviews instead of questionnaires to help the experts better understand the common understanding and our research goal.

Future work

We will further develop the UCD features such as role-based reporting and learning, system interoperability and instant communication, and will incorporate other UCD features which could improve reporting quality into the proposed system. Besides patient falls, more PSE subtypes, such as pressure ulcer, medication reconciliation will be supported by the system. The effectiveness and efficiency of UCD features will be initially evaluated through usability inspection and heuristic evaluation. Then, we will conduct population-based, individual-based, and group-based evaluation through user survey [24], interview and testing [25], and focus groups respectively.

Conclusion

We prototyped a user-centered PSE reporting system based on a PSE knowledge base, which includes PSE reports, solutions, and their connections. In this system, users can either choose an existing case or report a new PSE, then the system will retrieve similar cases and customized solutions based on the query and the reporter's role (e.g., manager, clinician, staff, patient). The user preference may be diverse for different purposes. The system allows the user to click the feedback button to indicate their preferences to a certain similar case or solution. All feedback will be returned to the algorithm implementation step in order to update the weights of similarity matrices and dynamically upgrade the system performance. This mechanism, similar to the ranking strategy of the Google search engine, will gradually stabilize the similarity matrices, making them more convincing as the feedback increases. It will be a win-win situation that both users and the system keep getting benefits from each other toward the common overarching goal of improving the PSE reporting quality and patient safety.

Acknowledgements

This project is supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), Agency for Healthcare Research & Quality (1R01HS022895), and University of Texas System Grants Program (#156374).

References

- [1] W. Runciman, P. Hibbert, R. Thomson, T. Van Der Schaaf, H. Sherman, and P. Lewalle, Towards an International Classification for Patient Safety: key concepts and terms, *Int J Qual Health Care* **21** (2009), 18-26.
- [2] P. Barach and S.D. Small, Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems, *BMJ* **320** (2000), 759-763.
- [3] J.T. James, A new, evidence-based estimate of patient harms associated with hospital care, *J Patient Saf* **9** (2013), 122-128.
- [4] M.A. Makary and M. Daniel, Medical error-the third leading cause of death in the US, *BMJ* **353** (2016), i2139.
- [5] A.K. Jha, N. Prasopa-Plaizier, I. Larizgoitia, D.W. Bates, and W.H.O.W.A.f.P.S. Research Priority Setting Working Group of the, Patient safety research: an overview of the global evidence, *Qual Saf Health Care* **19** (2010), 42-47.
- [6] Patient Safety Event Report, in, Center for Leadership, Innovation and Research in EMS, 2016.
- [7] Common Formats for Event Reporting - Hospital Version 2.0, in, The Agency for Healthcare Research and Quality.
- [8] T.K. Nuckols, D.S. Bell, S.M. Paddock, and L.H. Hilborne, Contributing factors identified by hospital incident report narratives, *Qual Saf Health Care* **17** (2008), 368-372.
- [9] H.J. Murff, V.L. Patel, G. Hripcsak, and D.W. Bates, Detecting adverse events for patient safety research: a review of current methodologies, *J Biomed Inform* **36** (2003), 131-143.
- [10] Y. Gong, Data consistency in a voluntary medical incident reporting system, *J Med Syst* **35** (2011), 609-615.
- [11] R. Shaw, F. Drever, H. Hughes, S. Osborn, and S. Williams, Adverse events and near miss reporting in the NHS, *Qual Saf Health Care* **14** (2005), 279-283.
- [12] Patient Safety Reporting System, in, NASA.
- [13] C. Kivlahan, W. Sangster, K. Nelson, J. Buddenbaum, and K. Lobenstein, Developing a comprehensive electronic adverse event reporting system in an academic health center, *Jt Comm J Qual Improv* **28** (2002), 583-594.
- [14] H. Kang and Y. Gong, A Novel Schema to Enhance Data Quality of Patient Safety Event Reports, in: *AMIA 2016*, Chicago, IL, 2016 (In press).
- [15] A. Sears and J.A. Jacko, *Human-Computer Interaction: Designing for Diverse Users and Domains*, CRC Press, Boca Raton, FL, 2009.
- [16] L. Hua and Y. Gong, Design of a user-centered voluntary patient safety reporting system: understanding the time and response variances by retrospective think-aloud protocols, *Stud Health Technol Inform* **192** (2013), 729-733.
- [17] Barriers to learning from incidents and accidents, in, European Safety, Reliability and Data Association, 2015.
- [18] K. Backstein, *The Blind Men and the Elephant*, Cartwheel, 1992.
- [19] WebM&M Cases & Commentaries, in, Agency for Healthcare Research and Quality.
- [20] Targeted Solutions Tool, in, Joint Commission Center for Transforming Healthcare.
- [21] M. Feil, Falls Prevention: Pennsylvania Hospitals Implementing Best Practices, *Pa Patient Saf Advis* **10** (2013), 117-124.
- [22] F. Healey and S. Scobie, The third report from the Patient Safety Observatory: Slips, trips and falls in hospital, in, National Patient Safety Agency, 2007.
- [23] L. Hua, S. Wang, and Y. Gong, Text prediction on structured data entry in healthcare: a two-group randomized usability study measuring the prediction impact on user performance, *Appl Clin Inform* **5** (2014), 249-263.
- [24] Y. Gong, H.Y. Song, X. Wu, and L. Hua, Identifying barriers and benefits of patient safety event reporting toward user-centered design, *Safety in Health* **1** (2015).
- [25] L. Hua and Y. Gong, Usability evaluation of a voluntary patient safety reporting system: Understanding the difference between predicted and observed time values by retrospective think-aloud protocols, *15th International Conference on Human-Computer Interaction, HCI International 2013* **8005 LNCS** (2013), 94-100.

Address for correspondence

Yang Gong, MD, PhD
7000 Fannin St Suite 600, Houston, TX 77030, USA
Yang.Gong@uth.tmc.edu
Phone: +1 713 500 3547, Fax: +1 713 500 3907.

OpenClinic GA Open Source Hospital Information System Enabled Universal Health Coverage Monitoring and Evaluation in Burundian Hospitals

Gustave Karara^{a, b}, Frank Verbeke^{a, b}, Evelyne Ndabaniwe^b,
Etienne Mugisho^c, Marc Nyssen^a

^a Department of Public Health, Biostatistics and Medical Informatics, Vrije Universiteit Brussel, Brussels, Belgium

^b Burundi Health Informatics Association, Bujumbura, Burundi

^c Belgium Technical Cooperation, Bujumbura, Burundi

Abstract

The Universal Health Coverage (UHC) is at the center of the 2030 Sustainable Development Goals agenda. In this study, the authors made an evaluation of the patient health coverage indicators in eight Burundian hospitals from 2011 to 2016. The relevant UHC indicators were calculated on the basis of patient administrative and health insurance data, collected via OpenClinic GA, an information and communication technology (ICT) supported health management information system (HMIS). The results show that the patient health services coverage rate was 70.8% for inpatients and 46.0% for outpatients. The patient health services payment rate as the proportion of total health service costs was above the 25% threshold recommended by WHO for inpatients (30.2%) and for outpatients (43.1%). The patient out-of-pocket payment was below the threshold of 180USD per patient per year for public hospitals. This study demonstrated the possibility to assess the degree of UHC in developing countries, by using routine data extracted automatically from the electronic HMIS.

Keywords

Universal Coverage; Insurance; Health; Health Expenditure; Developing Countries

Introduction

The Universal Health Coverage (UHC) has become at the center of 2015-2030 Agenda for Sustainable Development Goals (SDGs) and has gradually integrated into health policies of countries [1-6]. Under UHC framework, there would be no patient out-of-pocket payment (POOP) that exceeds a given level of affordability for the patient financial risk protection. According to WHO, people in developing countries should not spend in average of 25% or more of their total health expenditure and a maximum of 40% of 1.25USD per capita per day (i.e. 180USD per year) as POOP (set at zero for the poorest and most disadvantaged people) to avoid the impoverishment [7, 8, 21].

Some sub-Saharan Africa countries have made remarkable efforts to move towards UHC. In Ghana, a tax-funded national health insurance system covers 95% of diseases that affect Ghanaians, enabling financial protection and expanding coverage [6]. By implementing ambitious reforms that started in 2000, with the goal of UHC, Rwanda currently sustains one of the most elaborate health insurance schemes: the Community Based Health Insurance Scheme (CBHI), which covers over 90% of the population [9].

The Burundian government's effort to spread a scheme similar to CBHI preceded the Rwandan state-driven approach by a decade and a half, but was far less successful [10]. Currently, small CBHI plans cover less than 1% of the population specifically among local associations (farmers, bicycle drivers, etc.). Four types of health insurance plans are currently observed in Burundi:

- CAM insurance plan (*Carte d'Assurance Maladie*). This is a national program, with revenue collection and management at the community level similar to CBHI. According to the Ministry of Health, in 2014, 23% of the population had adhered to the CAM.
- MFP (*Mutuelle de Fonction Publique*) insurance plan provides health insurance for public employees. This public insurance plan covers 3% of the population [43].
- Private health insurance plans initiated by commercial health insurance companies in the formal sector. The classic private insurance plans (*ASCOMA, JUBILEE, SOCABU and SONAVIE*) cover almost 2% of population.
- Free health services insurance plan initiated by the Government for all children under 5 years and pregnant women.

According to the "Demographic and Health Survey" conducted in Burundi in 2010, 22.5% of the population reported to have at least one health insurance coverage scheme [11]. The health coverage in Burundi has increased (23-30%) substantially following the integration of "Free healthcare" policy for pregnant women and children under 5 years in 2006. The World Bank and WHO statistics (2014-2015) reported in Burundi a POOP of 4.39USD per year representing 20.9% of total expenditure on health [21, 22].

Monitoring health coverage indicators on UHC remains a challenge because the primary information of UHC comes from household surveys and health facility data but not health services data. Although most countries have functioning health facility-based health management information systems (HMIS), the HMIS data continue having a number of weaknesses, including incompleteness, inaccuracy and untimeliness, and therefore are not often used [12-14]. Appropriate application of Information and Communication Technology (ICT) can improve data quality by the computerization of the HMIS data [15, 16].

The introduction of open source ICT solutions for hospital management in several sub-Saharan health facilities prove that sub-Saharan countries move towards ICT development in health facilities [16, 17]. *OpenClinic GA* implementations are recorded in several health facilities over the world, and

monitored in more than fifty health facilities both public and private in sub-Saharan Africa [16, 18, 20]. *OpenClinic GA* is an open source integrated hospital information system developed by the project *ICT4Development* of *Vrije Universiteit Brussel* (VUB) and put in the public domain [19, 20]. The system covers management of administrative, financial and clinical patient records; lab, x-ray, and pharmacy data; and includes an extensive statistical and reporting module. *OpenClinic GA* was developed in Java connecting over JDBC to the most popular ANSI SQL 92 compliant database servers (such as MS SQL and MySQL Server). It offers an easy to use web interface facilitating HMIS deployment in often challenging technological settings commonly found in developing countries [16, 19].

This study attempts to show that UHC has been adequately evaluated in Burundian health facilities using *OpenClinic GA*-HMIS based on structured patient administrative and financial data (patient identification, type of encounter, insurance information, health services invoicing, etc.). The study analyzed data from eight hospitals among which district and national reference hospitals. Those hospitals joined the *OpenClinic GA* implementation projects since 2011. Most of them have been funded by the PAISS program (*Programme d'appui institutionnel au secteur de la santé*) of the Belgium Cooperation to start their ICT development [25].

Methods

The study was conducted during a 5-year period from 2011 to 2016. The process of *OpenClinic GA* implementation was applied and included (1) project management team set up, (2) *OpenClinic GA* software installation and configuration including security of the system, (3) users training and follow up; and (4) quality control, monitoring and evaluation. The implementation period was followed by a period of maintenance and assistance according to the needs of the hospital. The hospitals included in this study were:

- Four National reference hospitals (NRH): Military hospital of Kamenge (HMK, 2012), University teaching hospital of Kamenge (CHURK, 2014), Prince Louis Rwagasore clinic (CPLR, 2013) and Prince Regent Charles hospital (HPRC, 2015)
- Three District hospitals (DH): Ngozi regional hospital (NGORH), Muramvya (MUDH) and Kirundo (KIDH) district hospitals, all started implementation in 2015.
- One Private hospital (PH): Centre medico-chirurgical of Kinindo (CMCK, 2011)

We set up the *OpenClinic GA* software and configured the financial module by standardizing health insurance formats and health service components for all hospitals to facilitate the extraction of UHC indicators. We then collected and analyzed UHC-related data from the 8 hospitals in the period between 1/1/2013 and 30/06/2016.

The analysis of the collected patient information was performed in the *OpenClinic GA* statistics module. The pertinent indicators on UHC were centralized on the *Global Health Barometer* (GHB), a data warehouse installed on our servers at the VUB [20].

The most essential UHC indicators were:

- The patient health insurance coverage (PHIC) by evaluating patient health insurance data and the use of health insurance schemes in hospitals. We distinguished five types of health insurance schemes: (1) Free health services (FREE) where the patient did not pay anything, (2) Social health insurance (SHI) represented by the MFP plan, (3) Community based

health insurance (CBHI) including the CAM plan, (4) Private health insurance (PHI), and (5) No health insurance (PATIENT) where the patient paid the total of his health service expenditures.

- The patient health services coverage (PHSC) by evaluating patient's health services consumed and coverage of these services by health insurance schemes. We identified two categories of patients: (1) Insured patients for whom the POOP did not exceed 25% of the health services costs, and (2) Uninsured patients who covered 75% or more of the total consumed health services with POOP.
- The patient health services payments rate (PHSP) as the proportion of amounts paid by the patient for uncovered health services divided by total amounts of health services consumed.
- The patient out-of-pocket payment (POOP) as average amount paid directly by the patient for health services not (fully) covered by the health insurance scheme.

We separately calculated these metrics for out-patient and in-patient encounters. Comparative Chi Square testing was applied to compare the coverage rate of different health insurance schemes within the hospital and between different hospitals. The correlation analyses were applied to compare the distribution of health insurance schemes for out- and in-patient encounters and examine the relationship between insured and uninsured patient statuses. Finally, the UHC indicators' means were compared using the ANOVA test.

Results

Patient health insurance coverage (PHIC)

We analyzed more than 1.1 million electronic patient records in the course of our 4 years' study. The distribution of out and in-patients and their encounters is shown in table 1.

Table 1: Distribution of patients and encounters

Hospitals		Out-patients	In-patients	Out-patient encounters	In-patient encounters
NRH	HMK	586 905	48 487	734 260	48 581
	CPLR	113 146	18 043	215 196	19 139
	CHURK	126 042	25 939	208 717	45 789
	HPRC	50 395	12 287	77 969	77 969
DH	NGORH	27 454	13 452	51 471	18 700
	MUDH	10 164	2 691	14 311	3 095
	KIDH	9 778	5 469	16 197	6 053
PH	CMCK	82 228	2 619	90 807	2 619
Total		1 006 112	128 987	1 408 928	221 945

For each out- and in-patient encounter, part of the health service costs is paid by the health insurer and the remainder by the patient, according to the patient's health services coverage plan. We analyzed the health insurance schemes that have been used by patients for each encounter. Figures 1 shows the health insurance schemes coverage used for outpatient encounters in the eight hospitals.

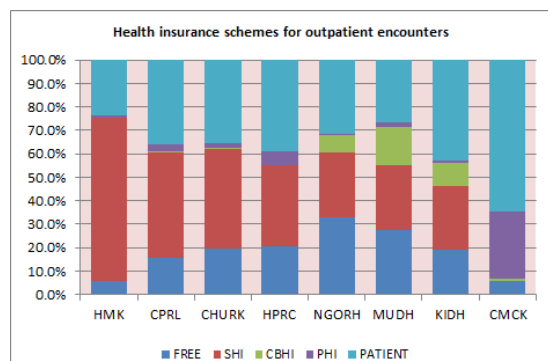


Figure 1 – Health insurance schemes for outpatient encounters

Free health services (FREE) and social health insurance (SHI) were the most frequently used schemes by outpatients. In district hospitals, FREE scheme was more frequently used ($p < 0.0001$) in 18.9% to 32.7% outpatient encounters. The SHI scheme was more frequently (34.3%–69.1%) used ($p < 0.0001$) in reference hospitals and especially in HMK (69.1%). The Community based health insurance (CBHI) scheme was mainly encountered (7.5%–16.1%) in district hospitals. This scheme was almost non-existent in other hospitals. The PATIENT scheme use was highest (64.8%) at the CMCK, the private hospital. In this hospital, private health insurance (PHI) plans were also mostly applied (28.4%) to outpatient encounters. For inpatient encounters, health insurance coverage schemes followed the similar distribution.

Patient health services coverage (PHSC)

Table 2 shows the health services coverage situation for out- and in-patients in the 8 hospitals during the study period.

Table 2 – Out- and In-patient health services coverage

Hospitals	Outpatient		Inpatient		Statistical significance (χ^2 test)
	Insured (POOP $\leq 25\%$)	Un-insured (POOP $\geq 75\%$)	Insured (difference)	Un-insured (difference)	
			69.2% (-8.5%)	23.6% (+6.6%)	$P < 0.001$
HMK	77.7%	17.0%			
CPRL	39.7%	39.9%	76.2% (+36.5%)	18.3% (-21.6%)	$P < 0.001$
CHURK	37.7%	43.0%	70.1% (+32.4%)	20.2% (-22.8%)	$P < 0.001$
HPRC	39.7%	47.0%	73.5% (+33.8%)	18.2% (-28.7%)	$P < 0.001$
NGORH	54.7%	34.8%	69.5% (+14.7%)	19.9% (-14.9%)	$P < 0.001$
MUDH	60.7%	22.0%	86.5% (+25.8%)	10.7% (-11.3%)	$P < 0.001$
KIDH	29.6%	48.4%	77.5% (+47.9%)	13.2% (-35.2%)	$P < 0.001$
CMCK	28.2%	70.3%	43.8% (+15.7%)	54.1% (-16.2%)	$P < 0.001$
Mean	46.0%	40.3%	70.8% (+24.8%)	22.3% (-18.0%)	$P < 0.001$
Statistical significance (ANOVA test)			$p = 0.005$	$p = 0.032$	

The proportion of inpatients insured was higher (+24.8%) than that for outpatients and this difference was statistical significant ($p = 0.005$). The difference observed on the PHSC in each hospital between in- and out-patients were statistically significant ($p < 0.001$). The overall picture of the PHSC in Burundian hospitals was that inpatients (70.8%) were covered better for health services than outpatients (46.0%).

Patient health services payment rate (PHSP)

The PHSP in the eight hospitals is represented in Table 3.

Table 3 - Out- and In-patient health services payment rate

Hospitals		Out-patients	In-patients	Difference
NRH	CHURK	40.5%	33.2%	-7.3%
	HPRC	41.2%	29.5%	-11.7%
	CPLR	46.3%	11.5%	-34.8%
	HMK	48.1%	56.7%	8.6%
DH	MUDH	25.2%	13.2%	-12.0%
	NGODH	36.8%	29.5%	-7.3%
	KIDH	47.4%	20.9%	-26.5%
PH	CMCK	59.3%	47.4%	-11.9%
Mean		43.1%	30.2%	-12.9%
Statistical significance (ANOVA test)				NS

The highest PHSP was found in the private hospital CMCK (59.3%) for outpatients and in the military hospital HMK (56.7%) for inpatients due to the high costs of health services in these hospitals that offer a high quality of healthcare compared to other hospitals in the country. The PHSP was normally lower in district hospitals (27.5%) than in national reference hospitals (35.4%). Averages of PHSP for outpatients (43.1%) and inpatients (30.1%) were above the threshold of 25% of the total amount of health services consumed by patients recommended by WHO.

Patient out-of-pocket payment (POOP)

The POOP in Burundian hospitals is shown in table 4.

Table 4 – Out and In-patient out-of-pocket payment

Hospitals		Out-patients	In-patients	Difference
NRH	CHURK	11.54 USD	98.98 USD	+758%
	HPRC	10.47 USD	93.68 USD	+795%
	CPLR	9.39 USD	40.51 USD	+332%
	HMK	10.84 USD	131.50 USD	+1113%
DH	MUDH	3.12 USD	2.10 USD	-33%
	NGODH	6.58 USD	24.71 USD	+276%
	KIDH	5.67 USD	12.22 USD	+115%
PH	CMCK	13.52 USD	393.42 USD	+2809%
Mean		8.89 USD	99.64 USD	+1021%
Statistical significance (ANOVA test)				n=0.064

The POOP's were higher in the private hospital CMCK and in the three national reference hospitals (HMK, CHURK and HPRC) than in district hospitals. They exceeded 90USD for inpatients in the first hospitals where health service tariffs applied were highest. Although the tariff of health services was almost the same in district hospitals of Burundi, the POOP was different in the 3 studied hospitals following to the level of health services coverage. The POOP average for inpatients (99.64USD) was significantly higher than the outpatient POOP (8.89USD). The largest differences of POOP between in- and outpatients were observed at CMCK and HMK. As seen above, the two hospitals offer health services that are expensive because of their (private) status and the quality of services they provide to the patient. The inpatient POOP's for the two hospitals were the highest and exceeded the threshold of 180USD for inpatients at CMCK.

Discussion

This study focused on the health services coverage monitoring by collecting and analysing data using *OpenClinic GA*-HMIS implemented in eight Burundian hospitals. The results showed that patient health services coverage (PHSC) was globally

70.8% for inpatients and 46.0% for outpatients. It was higher in the public hospitals compared to the private hospitals due to the important intervention of health insurance coverage plans oriented towards FREE (6.0%-55.6%) and SHI (5.8%-69.1%) schemes. We noted the intervention of CBHI (7.5%-26.0%) schemes in district hospitals especially for outpatient encounters. The CBHI scheme was predominantly based on the use of the CAM insurance plan. The results of health coverage found in the studied hospitals are higher than those in the reality at national level where the population health coverage is between 23-30%. The group of patients in the population is apparently better assured than the general population. This likely resulted from the adverse selection observed in certain health insurance contexts [23, 24] causing high costs for health insurance. Bearing in mind that the hospitals studied were at the second and third reference level, the PHSP remained globally above the 25%-threshold both for inpatients (30.2%) and for outpatients (43.1%) due to health services not covered by certain health insurance plans at those levels. This situation has been also observed in Rwanda [18, 24]. The POOP was also higher in these hospitals (8.89-99.64USD) than the national average (4.39USD) as could be expected. Although POOP was below the threshold of 180USD per year in all hospitals, it remained high for inpatients in private hospitals due to the high costs of health services in those health facilities. An effort is still needed for Burundian health insurance schemes to reach the patient financial risk protection in the framework of UHC.

Conclusions

In this study, we have demonstrated the feasibility of evaluating the UHC level in developing countries using ICT-HMIS routine patient data recorded by the health facilities themselves. Specifically, the implementation of an ICT-HMIS has enabled the monitoring and evaluation of UHC in eight hospitals of Burundi. The methods used allowed extraction of routinely collected patient data for secondary use in this health insurance coverage study. The study showed that the level of patient health services coverage was significantly higher for inpatients than outpatients. It was also high in hospitals where the patients were covered by more solidarity-based health insurance schemes especially by Free health services and social health insurance plans. We suggest that more efforts are needed to achieve good patient financial risk protection in Burundian hospitals since none of the hospitals involved in the study has reached all the thresholds recommended by WHO.

References

- [1] WHO, The World Health Report 2013, Research for Universal Health Coverage. Available from: http://www.who.int/universal_health_coverage/en/ [accessed 14 December 2016]
- [2] WHO, What is universal health coverage? Online Q&A, October 2012. Available from: http://www.who.int/features/qa/universal_health_coverage/en/ [accessed 14 December 2016]
- [3] WHA Resolution 58.33 2005. Social health insurance: sustainable health financing, universal coverage and social health insurance. Available from: http://apps.who.int/iris/bitstream/10665/20383/1/WHA58_33-en.pdf?ua=1 [accessed 14 December 2016]
- [4] UNDP, Sustainable Development Goals (SDGs), <http://www.undp.org/content/undp/en/home/sdgoverview/post-2015-development-agenda.html> [accessed 14 December 2016]
- [5] UNDP, SDGs, Goal3 targets, <http://www.un.org/sustainabledevelopment/health/> [accessed 14 December 2016]
- [6] Matshidiso M, Toward Universal Coverage in Africa, <http://www.afro.who.int/en/rdo/articles/4823-toward-universal-health-coverage-in-africa.html> [accessed 14 December 2016]
- [7] WHO & WB, Tracking universal health coverage, First global monitoring report, 2015. Available from: http://apps.who.int/iris/bitstream/10665/174536/1/9789241564977_eng.pdf [accessed 14 December 2016]
- [8] Boerma T, Abou Zahr C, Evans T, Monitoring Intervention Coverage in the Context of Universal Health Coverage, PLoS Med 11(9): e1001728 doi:10.1371/journal.pmed.1001728. Available from: <http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001728> [accessed 14 December 2016]
- [9] MoH, Rwanda Community based health Insurance policy, Kigali, April 2010. Available from: http://www.moh.gov.rw/fileadmin/templates/Docs/Mutual_policy_document_final1.pdf [accessed 16 December 2016]
- [10] Soors, W, Devadasan, N, Durairaj, V, Criel, B, Community Health Insurance and Universal Coverage: Multiple paths, many rivers to cross. World Health Report (2010). Background Paper, 48. Available from: http://www.who.int/healthsystems/topics/financing/healthreport/48_CHI.pdf [accessed 16 December 2016]
- [11] ISTEERBU, MFPDE, INSP, MSPLS, Enquête Démographique et de Santé (EDS 2010) au Burundi, Rapport Final Mai 2012. Available from: <https://dhsprogram.com/pubs/pdf/FR253/FR253.pdf> [accessed 16 December 2016]
- [12] WHO and Health Metrics Network (2011) Country health information systems. A review of the current situation and trends. Geneva: WHO. Available from: http://www.who.int/healthmetrics/news/chis_report.pdf?ua=1 [accessed 16 December 2016]
- [13] Aqil A, Lippeveld T, Hozumi D. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. Health Policy and Planning, 2009;1-12. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2670976/> [accessed 16 December 2016]
- [14] Health information systems in developing countries. A landscape analysis. California, Vital Wave Consulting, 2009. Available from: <http://www.minsa.gob.pe/geci/conferenciaops/recursos/43.pdf> [accessed 16 December 2016]
- [15] Simba O Daudi, Application of ICT in strengthening health information systems in developing countries in the wake of globalisation, Afr Health Sci. 2004 Dec; 4(3): 194-198. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688333/> [accessed 16 December 2016]
- [16] Verbeke F, ICT-based health information management methods in sub-Saharan hospitals, A field study in Rwanda, Burundi, Mali, Ivory Coast and the Democratic Republic of the Congo, Thesis, VUB-BISI 2012.
- [17] Braa J, Heywood A, Sahay S (2012) Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. Bull World Health Organ 90: 379-384. doi: 10.2471/blt.11.99580. Available from: <http://www.who.int/bulletin/volumes/90/5/11-099580.pdf> [accessed 16 December 2016]
- [18] Karara G, Frank V, Nyssen M, The Role of Hospital Information Systems in Universal Health Coverage Monitoring in Rwanda, Studies in health technology and informatics 08/2015; 216:193-7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26262037> [accessed 8 September 2016]
- [19] Verbeke F, OpenClinic GA. Available from: <https://sourceforge.net/projects/open-clinic/> [accessed 16 December 2016]
- [20] Vrije Universiteit Brussel, Globalhealthbarometer. Available from: <http://www.globalhealthbarometer.net/globalhealthbarometer/datacenter/datacenterHomePublic.jsp> [accessed 16 December 2016]
- [21] The world health report 2010. Health systems financing: the path to universal coverage. Geneva, World Health Organization, 2010. Available from: http://www.who.int/whr/2010/10_summary_en.pdf?ua=1 [accessed 16 December 2016]
- [22] World bank, 2014 World bank report. Available from: <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD/countries> [accessed 16 December 2016]
- [23] Parmar D et al. Adverse selection in a community-based health insurance scheme in rural Africa: Implications for introducing targeted subsidies. BMC Health Serv Res 2012;12:181. 21. Available from: <http://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-12-181> [accessed 21 December 2016]
- [24] WHO, Republic of Rwanda, Health financing system review, Options for universal coverage, Geneva 2008. Available from: http://www.who.int/health_financing/documents/hfsr_c_09-rwanda.pdf [accessed 21 December 2016]

- [25] BTC-CTB. Programme d'Appui Institutionnel au Secteur de la Santé (PAISS). Available from : <https://open.btcctb.org/fr/BDI/2083/p/programme-d-appui-institutionnel-au-secteur-de-la-sant-paiss.html> [accessed 21 December 2016]

Address for correspondence

Gustave Karara, MPH, PhD Candidate
Department of Public Health,
Biostatistics and Medical Informatics
Vrije Universiteit Brussel
Laarbeeklaan 103
B-1080 Brussels - Belgium
gustave.karara@vub.ac.be

User Centred Design and Nosocomials in Surgical ICUs: A Mobile Application for Peer Monitoring and Training in Hand Hygiene

Nadish Kariyawasam^a, Ming Chao Wong^a, Paul Turner^a

^a eHealth Services Research Group, University of Tasmania, Tasmania, Australia

Abstract

Nosocomial infections are a global public health risk. In low and middle-income countries the problem is acute with very high infection rates commonly contributing to poor patient outcomes including mortality. Organisational, cultural, and individual factors have been identified in these high rates, with poor hand hygiene compliance amongst clinicians a major risk factor. New approaches to achieving clinician behaviour change are required. User-centred approaches have proven effective to engage and support changes in clinician behaviours through the use of electronic tools. This paper reports on the experience of co-designing and implementing a mobile application with clinicians to enhance hand hygiene compliance. The peer monitoring and training supported by the application aims to directly contribute to evidence on reductions in infection rates in two surgical ICUs in Sri Lanka.

Keywords:

Mobile Applications; Cross Infection

Introduction

Nosocomial infections are a major public health concern throughout the world. Nosocomial infections contribute to elongation of hospital stay, long-term or permanent disability and death [1]. Each year health systems spend a considerable amount of resources, including high-end antibiotics, health professional work time, and hospital space to treat the consequences of nosocomial infections [2].

Previous research has already identified multiple factors contributing to differential rates of nosocomial infections in different clinical contexts. These include:

- Improper antibiotic usage and increasing antibiotic resistance;
- Congested and crowded healthcare facilities;
- Ineffective medical device cleaning and/or contamination during usage;
- Poor hand hygiene compliance amongst clinicians.

Numerous interventions have been implemented to reduce nosocomial infections with varying degrees of effectiveness. Infection rates, however, continue to be a significant burden on hospitals and patients, especially in lower and middle-income countries. Nosocomial infection rates in these countries continue to remain around 10 to 20 cases per 100 hospital admissions [3]. Sri Lanka, a 'low' middle-income country, continues to exhibit relatively higher rates of

nosocomial infections in the cases/100 admissions scale. Explanation of these higher rates forms part of the investigation, however, previous anecdotal evidences appear to be intimately related to organisational and cultural factors. It is also evident that within intensive care units (ICUs) where patients are more immobile and more immunocompromised, infection rates tend to be highest [4]. As a result, this research has focused on exploring mechanisms to contribute to reducing infection rates in Sri Lankan Surgical Intensive Care Units (SICUs).

Information technology in health care, (*eHealth*) has already opened up opportunities to increase health care safety and quality [5]. *eHealth* is also already extensively used to improve various situations associated with hospital-acquired morbidities. These include improving patient handover [6], reduce medical and medication errors [7], antimicrobial stewardship [8], electronic surveillance of nosocomial infections [9], and post-operative monitoring [10]. However, inadequate usage, over usage, and inappropriate usage of *eHealth* can also contribute unintended negative consequences. These include contributing to safety risks through technology-induced errors and medication errors [11].

In optimising the development, implementation, and impact of *eHealth*, human factors engineering and user centred design (UCD) approaches have been used extensively to ensure technology user attributes, needs and work contexts are adequately understood and addressed as part of information technology development life cycle [12,13]. In healthcare, UCD has contributed to emergence of *eHealth* systems that are versatile, user-friendly, safe and dependable and that integrate more seamlessly into clinical work environments [14,15].

This paper reports on the experience of co-designing and implementing a mobile application with ICU clinicians developed utilising user centred design with the aim of contributing to improving hand hygiene compliance in two hospital surgical ICUs in Sri Lanka. The peer monitoring and training supported by the mobile application aims to directly contribute to evidence on reductions nosocomial infection rates.

Methodology

Project Overview

This research involved using a multi-centred and multi-phased approach [16].

Phase one involved establishment of baseline quantitative nosocomial infection data across six surgical intensive care units in six Sri Lankan hospitals. Information from patient records were entered into an online data collection form that comprised data pertaining to demographic information, separation diagnosis, antibiotic usage, clinical interventions performed, culture reports as well as any other significant clinical findings indicative of infections.

In phase two, several observation sessions were carried out in two surgical intensive care units. During these sessions, hand hygiene compliance behaviours of unit staff were observed according to the 'five movements' of hand hygiene [17,18]. The observations were recorded using the World Health Organization, hand hygiene monitoring observation form. Semi-structured interviews and focus group discussions were carried out to determine staff attitudes and perceptions towards hand hygiene compliance and perceptions of nosocomial infection rates and 'best practice' for infection rate reduction. Staff were invited to consider whether they felt the current situation in their ICU should be changed and if so, what measures they would recommend to improve the situation. Staff were generally very supportive of trying to reduce infection rates and recommended the need for measures that would target improved hand hygiene compliance. Following further user-led discussions the idea of intermittent peer monitoring and iterative training in hand hygiene compliance was agreed and interest shown in how an eHealth system could be developed and deployed to support this desired behavioural change.

A mobile application was developed using rapid prototyping and iterative design with staff, based on their requirements and evolving feedback. The developed application was implemented in two surgical intensive care units as part of the commencement of phase three of the research methodology. The developed mobile application was implemented over a four month period in both ICUs. Alongside data collected by the application itself, data pertaining to behavioural and attitudinal changes associated with its implementation were also collected at the start of the second and fourth month of the application's implementation. Post-implementation observations and interviews/focus groups were conducted to explore perceptions around improvements in hygiene compliance rates, staff attitudes and perceptions towards hand hygiene compliance. The overall impact on compliance rates and perceptual and attitudinal changes will be compared against baseline data on completion of six months after application's implementation.

Phase four will involve further collection of patient record data pertaining to demographic, diagnosis, antibiotic usage; interventions performed, culture reports, and other clinical findings indicative of infections to contribute to baseline comparisons about the nature, type, and severity of nosocomial infections.

Utilising UCD in a mobile application for peer monitoring and training on hand hygiene to improve compliance

The design of the mobile application was an iterative process that involved co-design with, decision-making by and guidance from intensive care unit staff, hospital infection control staff, hospital administrators and microbiologists. At the initial stage of the research, users identified nosocomial infections as a major reason for the worsening of patient conditions while being treated. Then they identified leading causes of nosocomial infections including poor hand hygiene compliance. Hand hygiene was ultimately identified as one of the major contributing factors. Staff further anticipated that

proper intervention could significantly improve hand hygiene compliance. eHealth was identified as a potentially useful tool in the intervention. Staff suggested changes to the current 'Hand hygiene compliance observation form' and advocated incorporating this form into a more user-friendly and accurate mobile application for ease of access and use in the ICU environment and with staff workflow. A small tablet computer that could be easily cleaned with soft wipes was selected.

Development and Implementation of the Application

The mobile application was actually coded up using an agile development approach combined with iterative user feedback [19].

The application was based on the WHO hand hygiene compliance observation form but modified to accommodate local user needs and contexts of use. The iterative prototyping and user testing led to changing the app several times to reduce the 'number of clicks' and to improve the ease of 'repeated data entry'. The application collected hand hygiene compliance data during each peer monitoring session in each ICU and allowed time, staff member category, gender, and compliance with 'five movements of hand hygiene' to be easily and quickly recorded. This data was locally saved in the mobile device and synchronised to online system when connected to the internet.

Figure 1 shows the registration page of the developed application. New users registered their credentials while revisiting users can use their saved credentials to access the application.

Figure 1 – Registration screen of mobile application

Figure 2 shows the hand hygiene compliance monitoring screen of the application. Users can monitor up to four staff members at any given time. After multiple usability testing, the application was implemented in two surgical ICUs in two hospitals in Sri Lanka. The application is currently being used mainly by nursing staff to monitor fellow nurses and other staff. Unlike most of the other monitoring methods where one person is always the observer and the others are always observed, in this instance observer/observed iterate their roles as part of the training component around enhancing hand hygiene compliance. The staff were keen to ensure a supportive, learning and collaborative implementation rather than a more punitive, error-checking approach. Therefore, every participant who conducts observations will also at some other time themselves be monitored for compliance.

Figure 2 – Hand hygiene compliance data collection screen of the mobile application

During the initial phase of implementation, a decision was made to not provide any feedback on the observational findings. However, feedback is provided during the later part of the implementation to enable researchers to assess the impact of feedback on behavioural change. It is hoped this will contribute to enabling subsequent application roll-out to other hospital sites.

Results

Preliminary results of the research prior to provision of feedback are shown in tables 1-3. According to table 1, 12% of patients who were admitted to the SICUs had at least one positive culture report during their stay in the SICU or within 48 hours after leaving the SICU. However, regardless of the percentage of culture reports, 79% of patients who did not have any antibiotics prior to admission to SICU were started on antibiotics during their stay in SICU. Further, 60% of patients had signs, symptoms or investigations indicative of infections.

Table 2 demonstrates hand hygiene compliance rates observed by peers amongst a few categories of participant clinicians during the observations conducted prior to the provision of feedback. Based on these observations, nurses had a compliance rate of 42% while medical consultants and officers had a compliance rate of just above 20%. However, none of the electrocardiogram (ECG) technicians who visited SICU during observation sessions cleaned their hands.

Table 1– Percentage of patients with signs, symptoms, investigations and interventions indicative of nosocomial infections

Signs, Symptoms, Investigations, and Interventions	%
% of patients with a positive culture	12%
% of patients who were not on antibiotics at admission to ICU but started on antibiotics after admission	79%
% of patients with signs/ symptoms or investigations indicative of infections	60%

Table 2– Peer monitored hand hygiene compliance rates during first four months of implementation according to the service category

Service Category	Compliance rate
Medical Consultants	25%
Medical Officers	21%
Nurses	42%
Physiotherapists	31%
Radiographers	30%
ECG Technicians	00%

As reported in table 3, hand hygiene compliance rates are lowest during morning busy hours of 8 am to 12 noon closely followed by 12 noon to 4 pm. During the night-time hand hygiene compliance rates are relatively high with the highest rates reported during the hours of 8 pm and 12 midnight.

Table 3– Compliance rates during various times of the day

Hours of the day	Compliance rate
00 – 04	45%
04 – 08	49%
08 – 12	27%
12 – 16	28%
16 – 20	35%
20 – 24	50%

This research is currently in progress, and therefore the final outcomes are still unknown. However, several inferences can already be drawn from the available preliminary data:

1. Nosocomial infection rates appear to be high. However, when established filters are used to determine nosocomial infection rates using CDC guidelines the real infection rates found seems to be significantly lower than currently projected rates.
2. These discrepancies are mainly due to incomplete retrospective data hindering establishment of diagnoses.
3. Hand hygiene compliance is between 20% to 45% among health professionals. Compliance is comparatively high among Nurses while very low among ECG Technicians. Female professionals appear to comply better than males.
4. Out of five movements of hand hygiene, staff appear to be complying with proper hand hygiene measures with movement two and three (i.e., before aseptic procedures and after touching bodily fluids) but are least compliant with hand hygiene measures after movement five (i.e., after touching patients' surroundings).
5. Compliance rates are comparatively better during the night than when the ICU is busy during the daytime.
6. Almost all health care staff interviewed want an improvement in the situation to have better hand hygiene compliance and lower nosocomial infections. However, there seem to be huge differences in the level of commitment to change from individual to individual.
7. There is a marked difference between what health professionals perceive they are doing and their actual behaviour. Self-awareness of poor hand hygiene compliance seems to be lacking among most health professionals.
8. After implementation of the application, self-awareness of poor hand hygiene compliance has improved. Further, month to month hand hygiene compliance data appears to be on a positive trend.
9. Surgical Intensive Care Units staff seem to be more vigilant regarding potential introduction of infections from outside sources such as visiting staff and relatives of patients.

Discussion

This research is still in progress but provides a clear example of the benefit of utilising a user centred approach to ensure strong adoption and use of eHealth in a sustainable and inclusive manner. The UCD approach facilitated users to take more ownership of the application and contributed to improved awareness and support for the intervention. However, preliminary results have yet to indicate any marked improvement in compliance. Although quantitative analysis of how the changes and intervention have impacted on reducing nosocomial infections is yet to be done.

Further, this approach and ultimately the developed application help in supporting peer to peer mentoring among the health care workers, that is markedly different from Sri Lankan work culture which still exhibits significant hierarchy. This monitoring method may prove to be more beneficial in

inculcating self-awareness of the issue. Moreover, it is hoped the approach will assist in supporting sustained behavioural change compared to routine hand hygiene monitoring strategies assigned to infection control staff. Critically, it may be that the application enables a move away from an 'examination' approach to one that stimulates genuine and sustained behavioural change where participants can support one another to improve overall compliance.

In peer to peer observations, inculcating behavioural change happens in two ways. Firstly, staff tend to be more aware of complying when being observed. Then while they observe others, they identify how frequent compliance errors are and learn from others' mistakes and try to avoid them while they are at work. Regular feedbacks give the staff re-enforcement about good hand hygiene practice.

Together, this intervention may prove holistically to raise awareness of the existence of the problem, and provide staff with practical re-enforcement to improving behaviour.

Challenges around Nosocomials and Baseline Data

Pre-interventional hospital records are being analysed to provide a statistical baseline on nosocomial infection rates. In analysing these records it was initially found that relatively low nosocomial infection rates were recorded. However, this is primarily due to inability to formally establish genuine nosocomial infections due to the following reasons:

- Non-availability of appropriate culture reports conclusively validating the nature of any infection
- Poor recording of appropriate signs, symptoms, and investigations in the patient record that hinders ability to establish infections
- Non-availability of pre-infection reports inhibiting ability to definitively differentiate nosocomial infections from other types of infections
- Irrational antibiotic usage that inhibits use of some of the criteria conventionally deployed for formally establishing nosocomial infections

Given these limitations, the research team developed with a micro-biologist some criteria that could be used to get an indication of infection rates. In this context, a concept "Signs, Symptoms and Investigations Indicative of Nosocomial Infections (SSIINI)" was developed and utilised. The following criteria were used to consider a patient as having SSIINI:

1. No sign, symptom or investigation finding indicative of infections before admission to SICU or during first 48 hours at the SICU.
2. No record of antibiotic usage up to the first 48 hours at SICU.
3. A positive sign, symptom or investigation indicative of an infection first recorded at least 48 hours after admission to SICU.
4. Patients with signs, symptoms or investigations indicative of infections before spending 48 hours in the SICU or patients who have received antibiotics before spending 48 hours in the SICU (unless just one dose) were excluded from the analysis.
5. Patients who have spent less than 48 hours in SICU were also excluded from the analysis.

Limitations of research scope and scale

Quantitative analysis involving just six hospitals may not generate statistically significant data for generalisation of the results to all surgical ICUs. However, it is anticipated that any trends associated with the intervention may be used to predict the outcome of similar interventions in other locations.

This research is also being carried out as part of the doctoral thesis, i.e. this primarily involves one researcher. Therefore, there were some limitations in terms of finance, time, and trained personnel to conduct a larger and more comprehensive research investigation. However, adopting a multi-centred approach has contributed in excluding the possibility of any recorded outcomes caused purely by chance.

Conclusion

This research is contributing to efforts to reduce nosocomial infection rates. Using a UCD in co-designing a mobile application has led to strong and sustained adoption and use of an application perceived as user-friendly and efficient. It is anticipated that the research will also enhance understanding of how peer assisted monitoring can be used in inculcating behavioural change. Though final outcomes of the research are still unknown, preliminary insights show important steps towards change in user behaviour associated with this eHealth intervention. A future research paper will discuss the quantitative outcomes of the research and overall impact of the research in contributing to improved hand hygiene compliance and its relationship to overall reductions in nosocomial (SSI/NI) infections.

Acknowledgements

Authors wish to thank Ministry of Health Sri Lanka and all Administrative, Surgical Intensive Care Unit, Microbiology, Record Room and Infection Control staff for their participation assistance and guidance.

References

- [1] World Health Organization, Report on the Burden of Endemic Health Care-Associated Infection Worldwide, In, World Health Organization, Geneva, Switzerland, 2011.
- [2] E. Zimlichman, D. Henderson, O. Tamir, C. Franz, P. Song, C.K. Yamin, C. Keohane, C.R. Denham, and D.W. Bates, Health care-associated infections: a meta-analysis of costs and financial impact on the US health care system, *JAMA Intern Med* **173**(22) (2013), 2039-2046.
- [3] M. Meric, N. Baykara, S. Aksoy, I.O. Kol, G. Yilmaz, N. Beyazit, B. Mete, and H. Vahaboglu, Epidemiology and risk factors of intensive care unit-acquired infections: a prospective multicentre cohort study in a middle-income country, *Singapore Med J* **53**(4) (2012), 260-263.
- [4] A. Haddadi, M. Lemdani, and H. Hubert, Incidence, dependent and independent risk factors associated to nosocomial infections and to the mortality at the intensive care unit of the Timone University hospital, *Eur Sci J* **9**(18) (2013), 168-181.
- [5] D.W. Bates and A.A. Gawande, Improving safety with information technology, *N Engl J Med* **348**(25) (2003), 2526-2534.
- [6] M.C. Wong, E. Cummings, and P. Turner, User-centered design in clinical handover: exploring post-implementation outcomes for clinicians, *Stud Health Technol Inform*, **192** (2013), 253-257.
- [7] D.W. Bates, Using information technology to reduce rates of medication errors in hospitals, *BMJ* **320**(7237) (2000), 788-791.
- [8] J.C. McGregor and J.P. Furuno, Optimizing research methods used for the evaluation of antimicrobial stewardship programs, *Clin Infect Dis* **59** (Suppl 3) (2014), S185-192.
- [9] J.S. de Bruin, W. Seeling, and C. Schuh, Data use and effectiveness in electronic surveillance of healthcare associated infections in the 21st century: a systematic review, *J Am Med Inform Assoc* **21**(5) (2014), 942-951.
- [10] B. Debono, P. Bousquet, P. Sabatier, J.Y. Plas, J.P. Lescure, and O. Hamel, Postoperative monitoring with a mobile application after ambulatory lumbar discectomy: an effective tool for spine surgeons, *Eur Spine J* **25**(11) (2016), 3536-3542.
- [11] S. Redwood, A. Rajakumar, J. Hodson, and J.J. Coleman, Does the implementation of an electronic prescribing system create unintended medication errors? A study of the sociotechnical context through the analysis of reported medication incidents, *BMC Med Inform Decis Mak* **11**(1) (2011), 29.
- [12] M.C. Beuscart-Zéphir, E. Borycki, P. Carayon, M.W.M. Jaspers and S. Pelayo, Evolution of human factors research and studies of health information technologies: the role of patient safety, *Yearb Med Inform* **8**(1) (2013), 67-77.
- [13] G. Declercq, and X. Aimé, Why Medical Informatics (still) Needs Cognitive and Social Sciences, *Yearb Med Inform* **8**(1) (2013), 86-92.
- [14] O. Frykholm, M. Flink, M. Lindblad, and M. Ekstedt, User-centered design of integrated eHealth to improve patients' activation in transitional care, *Int J Integr Care* **16**(6) (2016), A338.
- [15] J. Chan, K.G. Shojania, A.C. Easty, and E.E. Etchells, Does user-centred design affect the efficiency, usability and safety of CPOE order sets?, *J Am Med Inform Assoc* **18**(3) (2011), 276-281.
- [16] N. Kariyawasam, M.C. Wong, P. Mahipala, and P. Turner, Reducing Nosocomial Infections: A User-centered Approach to Developing an eHealth system for Sri Lankan ICUs, In *Driving Quality in Informatics: Fulfilling the Promise* **208** (2015), 200-204.
- [17] H. Sax, B. Allegranzi, I. Uckay, E. Larson, J. Boyce, and D. Pittet, 'My five moments for hand hygiene': a user-centred design approach to understand, train, monitor and report hand hygiene, *J Hosp Infect* **67**(1) (2007), 9-21.
- [18] World Health Organization., *WHO guidelines on hand hygiene in health care : first global patient safety challenge : clean care is safer care*, World Health Organization, Patient Safety, Geneva, Switzerland, 2009.
- [19] K. Kautz, Investigating the design process: participatory design in agile software development, *Information Technology & People* **24**(3) (2011), 217-235.

Address for correspondence

Dr Nadish Kariyawasam
Nadish.Kariyawasam@utas.edu.au

An Experimental Comparison of a Co-Design Visualizing Personal Drug Information and Patient Information Leaflets: Usability Aspects

Soudabeh Khodambashi^a, Dagrun Haugland^a, Anette Ellingsberg^a,
Hanne Kottum^a, Janne Kutschera Sund^b, Øystein Nytrø^a

^a Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway

^b Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

Providing patients with specific information about their own drugs can reduce unintentional misuse and improve compliance. Searching for information is time-consuming when information is not personalized and is written using medical vocabulary that is difficult for patients to understand. In this study we explored patient information needs regarding visualizing of drug information and interrelationships by conducting a total of four co-design workshops with patients, other users and pharmacists. We developed a prototype and drug ontology to support reasoning about drug interactions. We evaluated individual performance in finding information, understanding the drug interactions, and learning from the provided information in the prototype compared to using patient information leaflets (PILs). We concluded that interactive visualization of drug information helps individuals find information about drugs, their side effects and interactions more quickly and correctly compared to using PILs. Our study is limited to co-morbid patients with transient ischaemic attack with several chronic diseases.

Keywords:

Drug Interactions; Patient Medication Knowledge; Computer Graphics

Introduction

Patients' misunderstanding about their prescribed medicines and their interaction can cause unintentional misuse, poor adherence and less effective treatment [1]. According to the pharmacists' association in Norway, over 1,000 patients die each year as a result of side effects and inappropriate use of drugs. Adverse consequences of the use of drugs are in addition to death or injury. Errors can occur because the patient is using medicines incorrectly due to insufficient information and miscommunication, and patients who are not in charge of their own medication are especially vulnerable to failure [2; 3]. Studies show that patients with limited literacy were more likely to misinterpret instructions, and precise wording on drug label instructions can improve patient comprehension [1]. On the other hand, according to the pharmacists' association in Norway, patients' poor compliance caused one out of three patients in Norway to not take their medicines due to fear of side effects. The number of patients with several chronic diseases is increasing and this has caused a decrease in the likelihood of good compliance due to the difficulty of keeping track of the reasons for taking drugs and how or when to take the medicine, especially by elderly patients [4; 5]. Studies show that better and more accessible information on drugs can help increase patient compliance [6].

There are many drug information sources available in Norway, including patient information leaflets (PILs) and websites for patients. Sources such as PILs, helsenorge.no, Legemiddel-håndboka.no and interaksjoner.no are available for drugs prescribed and marketed in Norway, however the available information is not individualized for each patient and often uses many medical words and phrases that might be difficult for individuals without knowledge in medicine to understand.

Although presenting and visualizing drug information to the healthcare professionals about adverse drug reaction, drug overdoses and drug combinations for multi-drug users has been a subject of interest for many years, enabling patients to easily access their drug information and be informed about adverse reactions has largely been neglected. Therefore, in this study we explored the individual's information needs regarding their medication, and how to visualize drug information (including drug interactions) in a way that is understandable for individuals without a medical background. Based on the collected data and involving patients and pharmacists, we developed a prototype that visualizes personal drug information. We evaluated individuals' performance in finding information, understanding the drug interactions and learning from the provided information in a prototype, compared to using PILs. To limit the scope, we focused on patients with transient ischaemic attack (TIA).

Material and Methods

In order to envisage concrete application examples, we developed personae and user stories in collaboration with a senior pharmacist to describe the target audience of the prototype. Use of personae in an interview with patients and users is a useful approach in order to avoid asking personal questions related to the interviewee's health condition and medication. The experiment was conducted in two parts. The goal of the first experiment was to identify requirements via co-design workshops in order to develop the prototype. The second experiment was performed to compare the visualization of drug information in the developed prototype to the PILs. In each of the experiments we applied different methods that are explained in this section.

Personae

In order to cover the largest possible number of common patient traits, a total of five personae were developed illustrating different types of patients. Three of the personae experienced transient ischemic attacks while suffering from other chronic diseases with the age range of 27- 87 years old. The other two personae were: one female persona with epilepsy was planning to become pregnant, the other persona

had multiple diseases (high cholesterol, diabetes, urinary tract infections) and therefore used number of other drugs with different symptoms such as headaches and depression.

Identifying Requirements through Co-Design Workshops

To identify requirements and explore how to visualize drug information, we applied co-design principles [7]. Hence, we conducted a total of four co-design workshops. Two of the co-design workshops were conducted together with pharmacists, and in the other two workshops we recruited participants (patients and volunteer participants) in order to include their perspectives on the design and visualizing of drug information. Results of the co-design workshops have been used to design and develop the digital prototype and competency questions [8] in order to design the drug ontology for the prototype. Details about the designated ontology based on the competency questions are provided in the “Design of the Ontology” section.

Co-Design Workshops with Pharmacists

A total of two pharmacists (one pharmacist who works at the counter at the hospital pharmacy and one clinical pharmacist) participated in our workshops. They were selected based on the purposive sampling method [9]. The goals of the workshops with the pharmacists were to understand patients' information needs about medicines, explore alternatives on presenting drug information to patients, identify functional requirements that the prototype needs to support, identify the competency questions that the ontology needs to answer, and create the conceptual design [10] of the prototype.

The first workshop consisted of five steps: 1) semi-structured interview on how pharmacists interact with patients; 2) semi-structured interview about what the pharmacists expected the patients' information needs to be; 3) pharmacists were asked to make a sketch of a system that would introduce personal drug information to a patient while thinking aloud; 4) discussion about the designed sketch from part 3; and 5) ranking the identified requirements based on their importance to be considered in the design of the prototype. To ensure the results were not affected by the order of the performed tasks, we reordered steps 2 and 3 for each pharmacist. The goal of the second workshop was to test the designed paper mock-up [10] and receive the pharmacists' feedback on the presentation and user interface.

Co-Design Workshops with Users

Since our project focused on patients with TIA, we contacted St Olav Hospital in Trondheim to recruit patients for our co-design workshops. A total of two patients diagnosed with TIA voluntarily participated in our case study. In addition, we recruited a total of ten students, employees and their families from the Norwegian University of Science and Technology (NTNU). Participation in our case study was voluntary, as we sent emails to student and employee mailing lists and requested volunteer participation. The goal of the workshops with patients and other volunteer participants was to identify drug information needs, elicit requirements for a user-centred design by involving users in the design of the prototype and to test the paper prototype. Furthermore, we also focused on identifying the competency questions that the ontology needed to answer in the workshops.

In the first user-workshop, the personae and user stories were given to the participants and we conducted a semi-structured interview in order to identify requirements with regard to the participants' information needs. In the second workshop, the designed paper mock-up was presented to the participants, and in the semi-structured interview they could give feedback on the presentation, including the detail of the presented

information, user interface and navigation, all based on the defined task set to evaluate whether all the tasks could be enacted. Results of the workshops contributed to the conceptual design of the prototype.

Design of the Ontology

To implement the prototype, we built a knowledge model of drug information in the form of an ontology. Competency questions define ontology requirements by indicating which questions the ontology should be able to answer with high-level coverage [8; 11; 12]. The competency questions are the questions that end-users would ask given the ontology about the drugs; hence in the four mentioned workshops we identified a set of competency questions.

We conducted a literature search to investigate the potential available ontologies. We identified Drug Ontology (DrOn) [13; 14], Drug-drug Interaction and Drug-Drug Interaction Evidence Ontology (DIDEO) [15], Drug Interaction Ontology (DIO) [16], Drug-Drug Interactions Ontology (DINTO) [17], and one comprehensive collection of marketed drugs in Japan, the USA and Europe (KEGG DRUG database) [18]. The ontology for our prototype was required to support reasoning and answer all the competency questions, and support Norwegian clinical terminologies (ordnett.no, ICD-10, MeSH terms, finnkode.no and Norsk Legemiddelhåndbok (the *Norwegian Drug Handbook*)). Since we could not address our requirements based on the identified ontologies and database, we designed our own ontology based on the personae, the competency questions and the *Norwegian Drug Handbook*. We note that since the project scope is limited, our ontology is only able to support medical terms and reasoning for the personae we developed for this project. Therefore, information about symptoms treated by a selected list of drugs related to personae (patients with TIA and multiple chronic diseases taking several drugs), active ingredients, adverse effects and dosage, are included in the ontology.

Comparison of Patient Information Leaflet and Prototype

We recruited a total of 13 participants by contacting St Olav Hospital and students at the NTNU. In order to reduce the risk of learnability of the scenario subjects, we divided the participants into two groups: one group (seven participants) only used the PILs while the other group (six participants) only had access to the developed prototype. We applied different methods to analyze and compare the PILs and the prototype that visualized the information about the drugs. The applied methods were: 1) pre-test questionnaire (questions about participants' age, their competence using a computer, and previous knowledge about the drugs we were questioning about in the case study to control for previous experience and its effect on the results); 2) scenarios and tasks (persona was given to participants followed by a total of seven tasks); 3) usability questionnaire (System Usability Scale (SUS) forms [19]); 4) semi-structured interview; 5) learning outcome questionnaire (to examine the extent to which the participant learned something from the given tasks without access to the prototype or PILs). In the last step (measuring learning outcome), we considered both ‘learning by remembering information’, which is linked to the given statements, and ‘learning that goes beyond the statements’. This was performed to investigate whether users acquire more knowledge using the prototype with intuitive visualization than with the PILs.

Data Analysis Methods and Metrics

For analyzing the collected data from semi-structured interviews and workshops, we applied the open coding method [20] in order to identify a list of concepts, group them based

on their similarities or related phenomena, explore the relationships between them and generate categories of concepts. We evaluated participants' task success rate by reviewing their written responses to the given questions in the scenario. We also measured the task completion time for each participant and calculated a geometric mean to present overall performance of the participants in task completion time, as the geometric mean is a better estimate for small samples ($n < 25$) [21].

Results

Information Needs

The interview with the pharmacists showed that they believed that all drug information was available on the Internet, but that some of the information was difficult to understand without the professional background of a pharmacist. The language, in which the information is to be presented, should be informal and easy to understand by the patient. As an example, none of the participants in the co-design workshops understood the term 'drug interaction'.

We identified a total of eight themes regarding the information needs: reason for taking a drug, dose and duration of treatment, practical use (when and how the drug should be taken, e.g. without food and with plenty of water), side effects, photos of the drugs, information about generic drugs, combination with vitamins supplements and natural products (herbal medicines), and their interaction. Based on the interview with the pharmacists, we found that information that is given by pharmacists to patients can be misunderstood. For example, one pharmacist told a story about a patient who was given an allergy spray because she was allergic to her cat. Later the same patient returned to the pharmacy because she was dissatisfied with the effect of the allergy spray. The patient explained that she had sprayed the cat every day, but that she was still suffering from allergic reactions. The patient had thus misunderstood and thought that the spray was for the cat and not to be used by her. Therefore, it is necessary to present the method of application or use of medicine and this should include over the counter drugs (OTCs). OTCs could also have side effects, interact with other drugs, including herbal supplements, and can sometimes cause serious health problems, especially in older adults [22; 23].

Visualizing Drug Information and Functionalities in Prototype

In the co-design workshops we explored how to present information about the drugs and visualize their interaction in textual description and graphs. We used the information about drug interaction from the web page at www.interaksjoner.no. However, we presented information using simple language than the web page and did not display any anatomic therapeutic chemical (ATC) codes. We also explored how to add additional features and information to the interaction graph, such as zooming, removable nodes, the ability to highlight specific parts of the graph, adding information about side effects, overlapping effects and interactions. We also explored presenting side effects in a word cloud, tiles inspired by VIE-VISU [24] and how to use the search function to find side effects. In the visualization we also evaluated presenting information to seniors who often have poor eyesight. Figure 1(a) presents our paper mock-up regarding drug interaction and Figure 1(b) presents the screenshot of the prototype.

The implemented functionalities, based on the co-design workshops and feedback received from pharmacists,



Figure 1– (a) Paper mock-up, (b) Prototype- in Norwegian

patients and users, are: login with high level of security (level 4) in Norway, which requires a national identity number, security token from a bank and a personal password [25]; add and remove drugs (including non-prescribed drugs) in the personal drug list; present possible side effects and indicate how common and how serious they are; present possible drug interactions and indicate how common and how serious they are; present what disease the drugs are intended to treat; present multiple drugs with the same active substances; display warnings for food and beverages that should not be combined with the drug(s); display warnings if the drug affects the ability to operate heavy machinery; present long-term effects of taking drugs; a warning to seek medical advice when necessary; include a photo of the drug; support to report experienced adverse reactions for each user; and display the expiry date of the drugs, customize how much information should be visible to them, and customize the display of information to specific populations (e.g. for elderly patients with poor eyesight). The prototype supports a search function, so it is also possible to search the drug symptoms either by entering the search keyword or by clicking on the body part of a human figure designated in the interface. The drug interactions and symptom interactions are presented in the graph in different colored clickable nodes with details presented in a pop-up window. The prototype supports the automatic updating of the interaction graph and the search results if a new drug is added to or removed from the drug list.

Comparison between Drug Information Leaflet and Prototype

People aged 23 to 70 participated in our experiment; the average age in group 1 (PILs) was 37.8 and in group 2 (prototype) 34.5 years old. Based on the pre-test questionnaire, all participants had good computer skills, their educational level was generally high, but was on average slightly higher in group 2. The term 'drug interaction' was unclear to the participants. Figure 2(a) presents a comparison of the geometric mean of task completion time between the two groups. Figure 2(b) presents the average SUS score between the two groups. Figure 3 presents the percentage of the task completion rate between the two groups.

We considered two different forms of learning in the evaluation (learning by remembering information linked to the statements and learning that goes beyond the statements). Figure 4 presents the percentage of learning outcome for group 1 and group 2. The learning outcomes were low for both groups, but we could see that users acquired some knowledge about their medicines, both through the PILs and the prototype; however, participants using the prototype seemed to learn more (learning by remembering information) about the drugs compared to the other group. Based on the presented comparison between the two groups, we could see that 'learning beyond the statements' was very low in both trials, but it was higher when using the PILs.

Based on the semi-structured interviews, participants stated that the PILs language was difficult to understand, lacked proper structure and contained too much information; it took quite a long time to read through the whole leaflet. They had to skim through the content rather than reading everything in detail. In addition, although some of the participants using the PILs could identify the drug interaction, they did not consider the statements to be correct as they did not understand the presented information. The prototype was described as intuitive, time-saving, easy and effective to use in finding information, logically structured and easy to understand. Several participants emphasized that the interaction graph made it easy to find information about drug interactions and how all the drugs worked together.

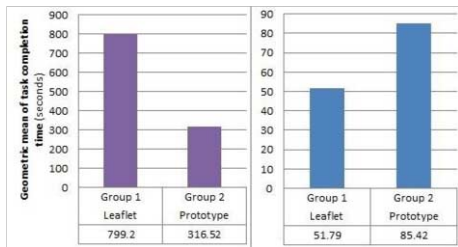


Figure 2– (a) Geometric mean of task completion time, (b) Right: average SUS score

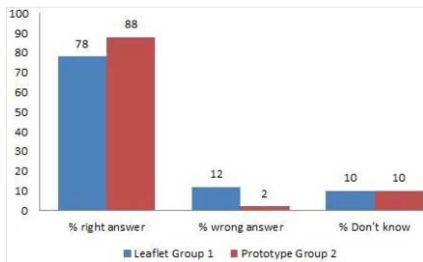


Figure 3– Comparison of task completion between two groups

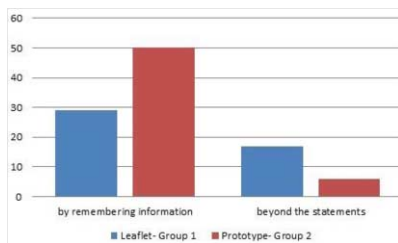


Figure 4– Comparison of learning outcome (in percentage)

Discussion

Based on the results in Figure 2, participants could finish the tasks faster when using the prototype than the PILs. We are aware that faster task completion may have led to more errors and affect the task success rate. Therefore, we compared the task completion time versus task success rate in order to analyze the time spent in relation to the number of correct answers. Figure 5 presents the average time spent for each correct answer. As is presented in Figure 5, the time spent (in seconds) was lower per correct answer using the prototype. Taking the participants' age and education into account and their effect on task completion time, we could not see any correlation between them.

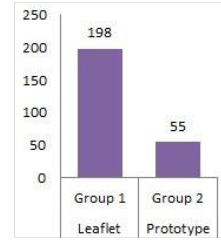


Figure 5– The average time spent (s) for each correct answer

We added adjectives to the SUS scores, as suggested by Bangor et al. [26], to better understand the SUS scores. The results of our SUS scores and their mapping to an adjective ranking scale are presented in Figure 6. We could see that the prototype is in the 'acceptable-excellent', while the PILs are in the 'marginal-low' range. This supports our findings from interviews and the results that are presented in Figure 2 and Figure 3, i.e. that the prototype is more usable than the PILs.

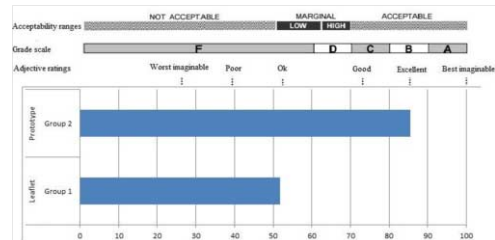


Figure 6– Adjective rating scale [26] added to SUS scores

One limitation of this study is that the sample is biased towards 13 participants, who could only access either the PILs or the prototype. Another limitation is that our prototype only supports certain personae and was limited to a certain number of drugs based on the developed personae. We did not evaluate the system in a clinical setting, and patients could have access to the system at home. This will be considered in future work. Conducting a case study in the future on both sources of information for all participants with a different set of tasks is another consideration. A new experiment may be carried out to explore how visualization could be adapted to different patient groups. For example, there is an opportunity to examine whether cancer patients have different needs to stroke patients, or whether women and men prefer different ways to view personal drug information.

Considering the results from the learning outcome, specifically 'learning that goes beyond the statements', a detailed study is required to investigate why users learned more from the PILs than the prototype, or to evaluate whether the results from our case study are generalizable or can be different with another set of participants.

Conclusion

In this paper we present the process of co-design to identify requirements for developing a prototype system for interactive and visualizing personal drug information. In addition, a case study was conducted to compare the developed prototype to PILs and evaluate users' task completion success rate, task completion time, learnability and usability. The analysis led to the assertion that using the prototype to find information is faster than using the PILs and the prototype is more usable than the PILs. Participants who used the prototype made fewer

mistakes. Understanding the drug interactions based on the PILs was cumbersome for the users. The results showed that presenting a lot of information in the PILs and using high-level language with no degree of personalization affected the task success rate and the time spent in finding answers. The experiment showed that, using the prototype users were more successful in recalling information related to the assessment of the statements. We conclude that the visual representation makes drug information more accessible and understandable to patients than PILs. The prototype is useful as it provides patients with the ability to assess their own drug use and to answer their concerns.

Acknowledgements

We gratefully acknowledge the support and feedback of Dr. Rune Sætre, Jens Lien and Marley-Kristin Singarajah.

References

- [1] T.C. Davis, A.D. Federman, P.F. Bass III, R.H. Jackson, M. Middlebrooks, R.M. Parker, and M.S. Wolf, Improving patient understanding of prescription drug label instructions, *Journal of General Internal Medicine* **24** (2009), 57-62.
- [2] S. Terje and J. Aarbakke, *Illustrated Pharmacology: Volume 1-Illustrert Farmakologi Bind 1*, Fagbokforl, 2010.
- [3] K. Bakken, E. Larsen, P. Lindberg, E. Rygh, and P. Hjortdahl, [Insufficient communication and information regarding patient medication in the primary healthcare], *Tidsskrift for den Norske lægeforening: tidsskrift for praktisk medicin, ny raekke* **127** (2007), 1766-1769.
- [4] J. Ulfvarson, C. Bardage, R.A.M. Wredling, C. VonBahr, and J. Adami, Adherence to drug treatment in association with how the patient perceives care and information on drugs, *Journal of clinical nursing* **16** (2007), 141-148.
- [5] R. Horne, J. Weinman, N. Barber, R. Elliott, M. Morgan, A. Cribb, and I. Kellar, Concordance, adherence and compliance in medicine taking, *London: NCCSDO* **2005** (2005), 40-46.
- [6] L. Osterberg and T. Blaschke, Adherence to medication, *New England Journal of Medicine* **353** (2005), 487-497.
- [7] E.B.-N. Sanders and P.J. Stappers, Co-creation and the new landscapes of design, *Co-design* **4** (2008), 5-18.
- [8] L. Rao, H. Reichgelt, and K.-M. Osei-Bryson, A Methodology for Developing High Quality Ontologies for Knowledge Management, in: *Knowledge Management for Development*, Springer, 2014, pp. 121-143.
- [9] M.D.C. Tongco, Purposive sampling as a tool for informant selection, *Ethnobotany Research & Applications* (2007), 12.
- [10] X. Ferré, N. Juristo, H. Windl, and L. Constantine, Usability basics for software developers, *IEEE software* **18** (2001), 22.
- [11] M.S. Fox, M. Barbuceanu, and M. Gruninger, An organisation ontology for enterprise modelling: preliminary concepts for linking structure and behaviour, in: *Enabling Technologies: Infrastructure for Collaborative Enterprises, 1995., Proceedings of the Fourth Workshop on, IEEE*, 1995, pp. 71-81.
- [12] A. Gangemi, Ontology design patterns for semantic web content, in: *International semantic web conference*, Springer, 2005, pp. 262-276.
- [13] J. Hanna, E. Joseph, M. Brochhausen, and W.R. Hogan, Building a drug ontology based on RxNorm and other sources, *Journal of biomedical semantics* **4** (2013), 1.
- [14] W.R. Hogan, J. Hanna, E. Joseph, and M. Brochhausen, Towards a Consistent and Scientifically Accurate Drug Ontology, in: *ICBO*, 2013, pp. 68-73.
- [15] M. Brochhausen, J. Schneider, D. Malone, P.E. Empey, W.R. Hogan, and R.D. Boyce, Towards a foundational representation of potential drug-drug interaction knowledge, in: *First International Workshop on Drug Interaction Knowledge Representation (DIKR-2014) at the International Conference on Biomedical Ontologies (ICBO 2014)*, 2014.
- [16] S. Yoshikawa, K. Satou, and A. Konagaya, Drug interaction ontology (DIO) for inferences of possible drug-drug interactions, *Medinfo* **11** (2004), 454-458.
- [17] M. Herrero-Zazo, I. Segura-Bedmar, J. Hastings, and P. Martínez, DINTO: using OWL ontologies and SWRL rules to infer drug-drug interactions and their mechanisms, *Journal of chemical information and modeling* **55** (2015), 1698-1707.
- [18] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, *Nucleic acids research* **38** (2010), D355-D360.
- [19] J. Brooke, SUS-A quick and dirty usability scale, *Usability evaluation in industry* **189** (1996), 4-7.
- [20] H. Priest, P. Roberts, and L. Woods, An overview of three different approaches to the interpretation of qualitative data. Part 1: Theoretical issues, *Nurse researcher* **10** (2002), 30-42.
- [21] J. Sauro and J.R. Lewis, Average task times in usability tests: what to report?, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2010, pp. 2347-2350.
- [22] J.H. Gurwitz, T.S. Field, L.R. Harrold, J. Rothschild, K. Debellis, A.C. Seger, C. Cadoret, L.S. Fish, L. Garber, and M. Kelleher, Incidence and preventability of adverse drug events among older persons in the ambulatory setting, *JAMA* **289** (2003), 1107-1116.
- [23] D.M. Qato, G.C. Alexander, R.M. Conti, M. Johnson, P. Schumm, and S.T. Lindau, Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States, *JAMA* **300** (2008), 2867-2878.
- [24] A. Rind, T.D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman, Interactive information visualization to explore and query electronic health records, *Foundations and Trends in Human-Computer Interaction* **5** (2011), 207-298.
- [25] DIFI, BankID-Agency for Public Management and eGovernment: <http://eid.difi.no/nb/bankid>, in.
- [26] A. Bangor, P. Kortum, and J. Miller, Determining what individual SUS scores mean: Adding an adjective rating scale, *Journal of Usability Studies* **4** (2009), 114-123.

Address for Correspondence

Soudabeh Khodambashi
Email: Soudabeh@idi.ntnu.no

Mutual Learning and Exchange of Health Informatics Experiences from Around the World - Evaluation of a Massive Open Online Course in eHealth

Sabine Koch, Maria Hägglund

Health Informatics Centre, LIME, Karolinska Institutet, Stockholm, Sweden

Abstract

We report our experiences from the Massive Open Online Course (MOOC), “eHealth – Opportunities and Challenges”, run by Karolinska Institutet using the edx platform both as session-based and self-paced versions between 2015 and 2016. In total, 13,302 students from 162 different countries were enrolled in our courses during the two-year period whereof 573 completed them. 331 students answered an exit survey after finishing the course which was analysed using quantitative and qualitative methods. As positive outcomes of the course, students highlighted set-up and content of the course, the pedagogical approach and the consistent international focus. Students lacked more practical case studies, more interactive discussions and proposed advanced follow-up courses on certain topics. Faculty lacked better functions for management of the discussion forum. Major advantages of the MOOC were mutual learning and exchange of health informatics experiences from around the world that would have been difficult to achieve in traditional learning contexts.

Keywords:

Education, Distance; Medical Informatics; Surveys and Questionnaires

Introduction

Although online education, virtual learning environments and distance learning platforms have been around for many years, it was not until 2008 that Dave Cormier coined the term Massive Open Online Course (MOOC) based on a large online course run by Siemens and Downes [1]. Massive Open Online Courses (MOOCs) offer new possibilities to reach large numbers of students with very different disciplinary and cultural backgrounds. Health informatics as a subject is interdisciplinary by nature. The challenges and opportunities of the field are global. Thus MOOCs may provide a good platform to address the global aspects of health informatics in an international environment.

Motivation to provide a MOOC

Since 2010 we run a global master's programme in health informatics at Karolinska Institutet as a joint collaboration with Stockholm University [2]. When Karolinska Institutet joined the edx consortium for MOOC-based education [3], we decided to design a basic course for a diverse target group with the aim to reuse parts of the course as a “flipped classroom” model for our own master's programme, but also

to market health informatics at Karolinska Institutet, to raise awareness for our education and to attract highly motivated and competent students for our campus education [4]. We ran different instances of the course, both as session-based and self-paced versions. Without any prior experience in providing MOOCs we offered a course that is broad in content and has an inhomogeneous target group. Thus we were interested in evaluating how students and teachers perceived the course and in how far our initial pedagogical considerations were valid.

Methods

Set-up and content of the course

“eHealth – Opportunities and Challenges” is a six-week introductory course in eHealth and health informatics targeting a broad student group without prior specific knowledge in the field [5]. Some experience with the health sector either as patient, care professional, IT and healthcare manager or policy maker is, however, recommended for following the course. The overall learning objectives of the course are to enable students 1) to describe different eHealth applications as well as enablers and barriers for their implementation; 2) to explain the importance of context, health informatics standards and terminologies for the design of eHealth applications and 3) to analyze eHealth strategies and discuss them in relation to the student's specific context.

The course is divided into six weeks, each covering a distinct topic (table 1). The content is released at the beginning of each week. Each week starts with an introductory video lecture about the topic of the week followed by a series of short videos (5-10 minutes) about sub-topics and by accompanying literature. Smaller learning tasks and quizzes are provided in each of the weeks. Every second week students have to complete a course assignment that covers the content of the preceding two weeks. The estimated workload is 4-6 hours per week. The course was given in three different versions:

1. As session-based course, provided free of charge without the possibility to opt for paid-for-certificates (April-June 2015).
2. As self-paced course (same content as 1 and free of charge) (Dec 2015 – May 2016)
3. As session-based course with some revised/added content, provided free of charge with the possibility to opt for paid-for-certificates (Oct – Dec 2016)

Table 1 - Set-up and content of the course

Week	Topic	Content	Special videos
Week 1	Introduction	<ul style="list-style-type: none"> - eHealth definitions and concepts - Types of eHealth applications - History of medical informatics - Opportunities and challenges of eHealth 	<ul style="list-style-type: none"> - Interview with an eHealth pioneer - Interviews with care professionals, IT developers, patients and informal carers
Week 2	eHealth for care professionals	<ul style="list-style-type: none"> - Health systems and healthcare organization - eHealth for care professionals 	<ul style="list-style-type: none"> - Interviews with different healthcare professionals in their respective work settings
Week 3	eHealth for patients and citizens	<ul style="list-style-type: none"> - ePatients and quantified-self - Ethical questions regarding ownership, access and use of data 	<ul style="list-style-type: none"> - Videos about concrete application example
Week 4	eHealth design	<ul style="list-style-type: none"> - Techniques for eHealth design - Importance of context and user involvement 	
Week 5	Technical prerequisites	<ul style="list-style-type: none"> - System architecture - Technical infrastructure - Standards and terminologies - Mobile eHealth 	<ul style="list-style-type: none"> - Use of a clinical scenario
Week 6	eHealth strategies	<ul style="list-style-type: none"> - Frameworks for set-up, analysis and implementation of eHealth strategies 	<ul style="list-style-type: none"> - Interviews with stakeholders from different countries

Pedagogical considerations

As described above, we saw the possibilities of MOOCs to provide a good platform to learn the global aspects of health informatics in an international environment. Targeting a varied student group the MOOC was sought to enhance inter-professional education (IPE) where students learn from, with and about each other, a practice that is lacking in many training programs [6]. Despite the advantages of IPE, the following challenges need to be considered: a) learning materials need to be presented in a congruent way; b) learning materials need to be adapted to the different levels of knowledge and experience of the students; and c) teachers need to enable each student to find her own learning track.

To tackle these challenges, we specifically considered two methods. First, we worked with video clips in the form of interviews with clinicians, industry representatives, policy makers, patients and informal carers from different parts of the world to bring practical experiences to the MOOC; and second, we worked with hand-drawn illustrations of patient scenarios to describe informatics challenges and technical prerequisites. Interactions with the students during the course were limited to one teaching assistant and the teacher responsible for a certain week answering questions in the discussion forum.

Evaluation methods

We distributed a link to an exit survey at the end of the course. The survey consisted mainly of structured questions and some open-ended questions. The structured questions were all based on a five point Likert scale and directed towards the following areas: Goal achievement; Value of different tools (videos, quizzes, assignments, and discussion forum) to achieve the learning outcomes; Common theme throughout the course; Skills development; Course satisfaction. Structured questions were answered with (To a very small extent (=1)...To a very large extent (=5)). Open-ended questions were related to

personal goals, course highlights and improvement suggestions. Open-ended questions and entries in the discussion forum related to opinions about the course were analyzed by inductive thematic content analysis. The questions resemble standard questions used for all course evaluations at our university and are used for continuous quality improvement and comparability between courses.

Results

Course participants

In total, 13302 students had enrolled in the courses and 573 (4.31%) students completed one of the course instances. Most students were between 26 to 40 years old with a median student age of 32 years. 40.7% were female, 59.3% male. Most of them were highly educated with 49.8% having an advanced degree. Students came from 162 different countries whereof the most represented countries were the United States 15.9%, India 9.7% and UK 4%. Course participants had varying backgrounds. 44.6% had a healthcare background, 24.1% had a background in computer science, 7.9% were health informaticians, 1.9% policy or decision makers, 1.6% patients or representatives for a patient association and 1.2% informal care givers. Students' personal learning goals varied. Most of them had a personal and professional interest in the topic and wanted to broaden their knowledge about eHealth. Figure 1 gives an overview over the participants in the respective courses.

	Session-based Course 2015	Self-paced Course 2016	Session-based Course 2016
Duration	6 weeks	4 months	6 weeks
# Participants	5 329 40% female, 60% male Median age: 32 51% have an advanced degree	5 239 42% female, 58% male Median age: 31 47% have an advanced degree	2 734 40% female, 60% male Median age: 33 51.4% have an advanced degree
Background	46.6% Healthcare 25.7% Computer Science 7.9% Health Informatics 2.5% Policy/Decision makers 1.9% Patients or representatives for a patient association 1.4% Informal care givers	46.0% Healthcare 24.6% Computer Science 8.8% Health Informatics 1.2% Policy/Decision makers 1.0% Patients or representatives for a patient association 1.6% Informal care givers	41.2% Healthcare 22.0% Computer Science 6.9% Health Informatics 10.0% Entrepreneurs 2.1% Policy/Decision makers 1.9% Patients or representatives for a patient association 0.6% Informal care givers
Countries	160 countries (U.S. 18.3%, India 11.1%, Sweden 5.1%)	162 countries (U.S. 15%, India 11%, UK 4%)	134 countries (U.S. 14.3%, India 7.1%, UK 4%)
Completed course with honor certificate	342 (6.42%)	145 (2.77%)	86 (3.15%) completed ; 36 (of 70 paid certificate)

Figure 1 – Background of course participants

Quantitative results from the exit survey

331 students filled in the exit survey. 71.6% (n=237) perceived that they achieved all learning outcomes of the course to a large or very large extent. 56.1% (n=186) reported that they achieved their personal goals for taking the course to a large or very large extent.

74.9% (n=248) found to a large or very large extent that there was a common theme running throughout the course. 49.5% (n=164) said that they developed valuable expertise and skills to a large or very large extent.

Regarding the different learning activities during the course students enjoyed the videos most. 83.7% (n=277) found them valuable or very valuable, followed by the quizzes (78.2%; n=259) and the course assignments (73.1%; n=242). The discussion forum was considered valuable or very valuable by only 37.2% (n=123) of the students. 69.8% (n=231) were satisfied or very satisfied with the support offered by the staff and teaching assistants.

87.3% (n=289) agreed or strongly agreed to recommend the course to other students.

Table 2 gives a comparative overview over the quantitative results between the three course instances.

Table 2 - Quantitative data analysis results (weighted average of Likert scale (1-5 where 5 is best))

Question	First MOOC (n=184)	Self-paced MOOC (n=64)	Revised MOOC (n=78)
In my view, I have achieved all the learning outcomes of the course	3.9	3.8	3.9
In my view, I have reached my personal goal for taking the course	3.9	3.6	3.5
In my view, there was a common theme running throughout the course	4.0	4.0	4.0

In my view, I developed valuable expertise and skills	3.6	3.3	3.7
How valuable were the videos to help you reach the learning outcomes?	4.4	4.3	4.3
How valuable were the quizzes to help you reach the learning outcomes?	4.1	4.0	4.2
How valuable were the course assignments to help you reach the learning outcomes?	4.1	3.8	4.0
How valuable was the discussion forum to help you reach the learning outcomes?	3.2	3.1	3.0
How satisfied were you with the support offered by the staff and teaching assistants?	3.9	4.4	3.8
I would recommend this course to others	4.4	4.1	4.3

Qualitative results

What did students enjoy most?

Analysis of the open-ended questions of the exit survey, discussion forum entries and open course evaluations on the Internet resulted in the following benefit categories: Comprehensiveness, International Focus and Pedagogical Approach (table 2). Students highlighted set-up and content of the course that was considered to give a holistic, worldwide overview over the eHealth landscape. The global, multidisciplinary perspective of the course and the consistent worldwide aspects in the material were acknowledged by many students. The course was considered very educational and easy to understand. Especially the inclusion of practical knowledge provided through numerous interviews with different stakeholders was highly appreciated.

What can be done better?

Improvement suggestions were related to Content and Interaction and Feedback (table 2). Students lacked more practical examples in form of case studies, requested further deepening of technical issues as well as eHealth implementation and proposed more advanced follow-up courses on these topics. One student also proposed to provide a complete online master program in MOOC format. Students further asked for more interaction in the discussion forums and feedback on their assignments. They proposed less multiple choice quizzes and addition of online conversation.

Table 3 - Qualitative data analysis results

Theme	Category	Category description	Example quotes
Benefits	Comprehensiveness	In this category remarks and comments regarding the inclusiveness and understandability of the content as well as the structure of the course were included.	<p>"It [the course] is covering the HIT landscape of whole of the world. I am a volunteer with the state government of [an Asian country] on e-health projects and I have recommended this MOOC to the team for its sheer brilliance and comprehensiveness. Especially the resources and the timeline in which they have been structured. This MOOC is the perfect example of congruence."</p> <p>"I have been very energized by going through the eHealth online course. From doing piecemeal work on individual eHealth modules starting 10 years ago, we have switched to preparing for the next generation electronic health record system and have gathered a multi-disciplinary team of medical informaticians spending up to 50% of our time in this field. I have encouraged my team members to enroll in this ehealth course and learn together."</p> <p>"It [the course] glued and streamlined different concepts in the perfect big picture."</p> <p>"I have now the language and knowledge to discuss quality assurance issues at work with senior managers."</p>
	International Focus	In this category remarks and comments concerning the focus on international and global contexts were included.	<p>"I liked the approach of teaching eHealth from different, global perspectives."</p> <p>"The course showed the power of MOOC not only to reach a lot of students but also get their contributions."</p>
	Pedagogical approach	This category includes remarks and comments regarding the use of interviews, scenario descriptions or other ways of presenting the content.	<p>"The course was very insightful and above all innovative in its presentation by incorporating experiences and challenges from real professionals into the course lectures. This gave a very good contextual richness to the materials taught."</p> <p>"Care professionals' experiences add great insight into learning this course. It brings the lectures to life and places topics discussed into proper context and perspectives. Thanks for the innovation in the course presentation."</p> <p>"Sufficient care was taken to present lectures in a way that even students like myself from a different cultural background could grasp concepts easily."</p>
Areas for improvement	Content	This category includes remarks and comments regarding additional content.	<p>"It would be great to have practical case studies from different countries, more examples of IT systems used including screen-dumps."</p> <p>"Add more references about implementation support and integration guidance in the second version of the course."</p> <p>"I noted that in Healthcare Informatics Standards FHIR was only named. I strongly recommend one whole week for FHIR standards."</p> <p>"Please develop an advanced level to this course offering."</p> <p>"I propose an advanced course where computer skills are a pre-requisite."</p>
	Interaction and Feedback	In this category comments and remarks about the interaction between students, students and teachers and feedback from teachers or teaching assistants.	<p>"The discussion forum was crowded. Course leaders could identify entries each week and feedback on them."</p> <p>"Provide a mobile App for course interaction."</p> <p>"The addition of a peer graded assignment would be an interesting consideration for future courses."</p>

Discussion

We exploited the advantages of a diverse, international student group by highlighting similarities and differences between countries. Many students were inspired by this approach. As an example, one of the most discussed points in the discussion forum in our first course dealt with the question whether electricity is a prerequisite for eHealth or not. Many

participants had not even thought about this being a major limitation in low income countries. As also highlighted in the exit survey, interactivity between students and between students and teachers was considered to be an area for improvement. Interestingly, there were no major differences in students' perceptions between the session-based courses vs. the self-paced one (table 2), except in regard to satisfaction with staff support. Staff support was evaluated better in the self-paced course which might relate to the fact that it was

clearly stated in the instructions that support was not offered. So students were happy when getting some support anyway. Chen et al consider student-faculty interaction and peer-to-peer collaboration essential parts in student engagement which is positively related to the quality of the learning experience [7]. Experiences from this course are that the current edx platform does not sufficiently support such student engaging activities, which is in line with criticism towards xMOOCs [8]. Although the discussion fora were meant to facilitate student-faculty interaction, their cumbersome design hindered this. If the learning platform in use does not meet these requirements, we suggest complementing with tools that support more student-faculty interaction as well as peer-to-peer collaboration. An alternative approach would be to use more interactive online tools, e.g. a combination of live broadcasting via periscope with live chat possibilities in twitter. Online learning is not only a question of having access to content; rather, it should also provide opportunity for social learning through interaction and connectedness [9]. If the learning platform in use does not meet these requirements, we suggest complementing with tools that support more student-faculty interaction as well as peer-to-peer collaboration. Also the integration of tools for adaptive eLearning would be valuable to adapt the learning to individual students' profiles and behaviors [10].

Today, videos are reused for classroom teaching. Its' potential impact on student recruitment for our Master's program is too early to state as we only could monitor one round of admissions so far.

Possible future directions could be to reuse the course material as Small Private Online Course (SPOC) for continuing education and to develop more in-depth courses for certain topics [11]. A major drawback is however the lack of incentives in our current reimbursement system for MOOCs.

Conclusions

Experiences from our MOOC with participants from 162 countries highlighted both challenges and benefits. A difficulty encountered by both students and staff during the course was the poorly designed discussion forum which affected the interaction in the course negatively. A major advantage of the MOOC was the mutual learning and exchange of health informatics experiences from around the world – a learning that would have been difficult to achieve in traditional learning contexts.

Acknowledgements

We thank all guest lecturers, interview persons and teaching assistants as listed in the biography of course staff [5]. The course production was supported by the Centre for Learning and Knowledge at Karolinska Institutet. The rebranding of the third version of the MOOC was funded by EIT Health (www.eithealth.eu).

References

- [1] Cormier D, Siemens G. Through the open door: open courses as research, learning, and engagement. *EDUCAUSE Review*. 2010; 45(4): 30-9
- [2] Joint Master's Programme in Health Informatics, Karolinska Institutet, Stockholm, Sweden. Available at <http://ki.se/en/utbildning/5hi12-masters-programme-in-health-informatics> (last accessed April 11, 2017).
- [3] edx Consortium. Available at <https://www.edx.org/> (last accessed April 11, 2017).

- [4] Alvarez B. Flipping the classroom: Homework in class, lessons at home. *Education Digest: Essential Readings Condensed For Quick Review*. 2011; 77(8): 18-21.
- [5] eHealth – Opportunities and Challenges. MOOC delivered by Karolinska Institutet. Available at <https://www.edx.org/course/ehealth-opportunities-challenges-kix-kiehealthx#!> (last accessed April 11, 2017).
- [6] Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. 2010; 376(9756): 1923-58
- [7] Chen PSD, Gonyea R, Kuh G. Learning at a Distance: Engaged or Not?. *Innovate: Journal of Online Education* 2008; 4(3)
- [8] Vardi MY. Will MOOCs destroy academia? *Commun. ACM* 2012 Nov 01;55(11):5.
- [9] Brindley JE, Walti C, Blaschke LM. Creating Effective Collaborative Learning Groups in an Online Environment. *The International Review of Research in Open and Distributed Learning*. 2009; 10(3)
- [10] Shute V, Towle B. Adaptive e-learning. *Educational Psychologist*. 2003; 38(2): 105-114.
- [11] Fox A. From MOOCs to SPOCs. *Communications of the ACM*. 2013; 56(12): 38-40.

Address for correspondence

Sabine Koch, PhD, FACMI
 Professor and Director
 Health Informatics Centre
 Department of Learning, Informatics, Management and Ethics
 Karolinska Institutet
 171 77 Stockholm
 SWEDEN
 e-mail: sabine.koch@ki.se

50th Anniversary International Medical Informatics Association (IMIA) History Working Group and Its Projects

Casimir A. Kulikowski^a, George Mihalas^b, Robert Greenes^c, Valerio Yacubsohn^d, Hyeoun-Ae Park^e

^a Department of Computer Science, Rutgers-The State University, Piscataway, New Jersey, USA

^b Victor Babes University of Medicine & Pharmacy, Timisoara, Romania

^c Department of Biomedical Informatics, Arizona State University, Tempe, Arizona, USA

^d School of Public Health, Faculty of Medicine, University of Buenos Aires, Buenos Aires, Argentina

^e College of Nursing, Seoul National University, Seoul, Korea

Abstract

The IMIA History Working Group has as its first goal the editing of a volume of contributions from pioneers and leaders in the field of biomedical and health informatics (BMHI) to commemorate the 50th anniversary of IMIA's predecessor IFIP-TC4. This paper describes how the IMIA History WG evolved from an earlier Taskforce, and has focused on producing the edited book of original contributions. We describe its proposed outline of objectives for the personal stories, and national and regional society narratives, together with some comments on the evolution of Medinfo meeting contributions over the years, to provide a reference source for the early motivations of the scientific, clinical, educational, and professional changes that have influenced the historical course of our field.

Keywords:

Medical Informatics; Nursing Informatics; Organizations

Introduction

The 16th World Congress of Medical Informatics - Medinfo 2017 in China, will mark the 50th Anniversary since the International Federation of Information Processing (IFIP) Societies approved the formation of the Technical Committee (TC) 4 on Medical Information Processing, which was the predecessor of IMIA. The IMIA History Working Group (WG), since 2014, has been working on organizing and writing about the international history of biomedical and health informatics (BMHI), and is editing a book of contributed articles from leaders in the field to commemorate the occasion. This article provides an overview of the work leading up to the book, and describes a preview of the goals, structure and evolution of what it is intended to cover.

Methods

The IMIA History eBook to be published by IOS Press, will contain original autobiographical retrospectives by pioneers and leaders in the field, together with professional organizational histories of the national and regional societies and working groups of IMIA, as well as interpretive commentary on a number of important themes and topics, which have evolved changes as scientific and clinical practices under the influence of new insights, technologies, and the changing socio-economic, cultural and professional circumstances around the globe over the past 50 years.

IMIA History Taskforce

Discussions during IMIA Board meetings during the first decade of the millennium brought up the subject of recording more systematically the history of the organization, and how the structure and functions evolved with changing research and practice trends in the field. A Taskforce on the History of Biomedical and Health Informatics was proposed and approved in 2009, and was composed of representatives from all the major regional associations: Casimir Kulikowski (USA/AMIA-North America, Chair), George Mihalas (Romania/EFMI-Europe), Hyeoun-Ae Park (Korea/APAMI – Asia-Pacific), Sedick Issacs (South Africa/HELINA – Africa), Alvaro Margolis (Uruguay – Latin America). The long-time secretary of IMIA, Diarmuid UaConnail, was able to provide many original documents of the Association, including minutes from past Board and General Assembly meetings. A Wiki media repository was set up at Rutgers University (<http://infohistory.rutgers.com>) by Charles McGrew to make these documents openly accessible, and to index them and summarize the time-lines of IMIA developments graphically, together with thumbnail sketches of past leaders of IMIA. These materials are aimed at chronicling the development and evolution of IMIA, its contributors, its sponsored events, its publications, and its history. The role of the Taskforce was to coordinate and channel the interests of IMIA members related to the history of the organization, to lead the recording of personal and group recollections, and to write about how the field has developed from a wide range of international perspectives. The members of the Taskforce solicited articles for the IMIA Yearbook about the History of the field, and also contributed their own perspectives [1,2,3,4,5,6,7].

Over the past years, the IMIA History Taskforce organized workshops and panels at many conferences, including: Medinfo 2010 in Cape Town, the IMIA-LAC 2011 meeting in Guadalajara [8], many AMIA Annual Meetings and ACMI Meetings since that date, the Oslo MIE 2011, APAMI in Beijing in 2012, the Prague Workshop held in conjunction with the EFMI-STC in April 2013, the Medinfo 2013 Congress held in Copenhagen [9], MIE 2014 in Istanbul [10], and APAMI 2014 in New Delhi. A related historical anniversary took place in 2011 in Heidelberg – the 50th year of publication of the first and official IMIA journal of medical informatics: *Methods of Information in Medicine (MIM)*. The IMIA President, Reinhold Haux organized a symposium, which discussed many changes that have taken place in medical and health informatics since the founding of *MIM* by Gustav Wagner in 1962. The Copenhagen Medinfo 2013 saw important discus-

sions on the systematic elicitation of articles on the history of the field, while at the Prague EFMI-STC, earlier that same year, there had been a special workshop on European History, which led to publications in the proceedings [10, 11, 12, 13] and later a set of books edited by two of the participants, Izet Masic and George Mihalas [14], who compiled papers and biographical materials. In order to recognize the activities of the IMIA History Taskforce, its members proposed to transform themselves into a Working Group, and this was approved by the IMIA General Assembly in August 2014 at the APAMI New Delhi meeting.

IMIA History Working Group

Shortly after the IMIA History WG was established, there was an important anniversary in the USA – the 30th anniversary of the American College of Medical Informatics (ACMI). Kulikowski, who also chairs the ACMI Committee of Historians, worked with Alexa McCray to organize two commemorative panels [15] at AMIA 2014: the first with pioneers who were instrumental in the formation of ACMI, and the second with leaders of the field to cover the three decades intervening. After this, the IMIA WG prepared a workshop for Medinfo 2015 in Sao Paulo, where the plans for the 50th Anniversary book were discussed, and prospective contributors identified and invited. At the same time, plans were made to participate in the EFMI-STC to be held in Paris the following April of 2016. This was important because included a presentation on the founder of IFIP-TC4, Dr. Francois Gremy, by his daughter Isabelle, and another on Dr. Peter Reichertz, who was a most influential pioneering researcher and teacher in medical informatics from Hannover in Germany, by Rolf Engelbrecht. In the remainder of 2016, three conferences were instrumental in obtaining further commitments from contributors to the proposed History Book from around the world – Nursing Informatics 2016 in Geneva, HEC-MIE 2016 in Munich, and the AMIA 2016 meeting in Chicago. At EFMI, the IMIA General Assembly voted to fund the publication costs of the IMIA History, thus making it possible to proceed with formal invitations and plans for publication with IOS Press. At the AMIA 2016 meeting, there was yet another anniversary to celebrate – the Symposium of Computers in Medical Care (SCAMC), which was the main vehicle for disseminating research in medical informatics in the USA before the creation of AMIA. The meeting brought together leaders in writing the history of nursing, public health, and other clinical informatics to help coordinate their contributions to the IMIA History.

Some Historical Highlights of IMIA-Related Biomedical and Health Informatics in Asia

The history of biomedical and health informatics in Asia goes back to the establishment of several national societies, and the decision by IMIA to hold the Third World Congress on Medical Informatics - MEDINFO-80 [17] under the leadership of Dr. Shigekoto Kaihara, a pioneer in medical hospital systems and computer-based decision-making at Tokyo University, who also established the Japan Association for Medical Informatics (JAMI). Many international and Asian participants came to this first Asian MEDINFO, including delegates from China, establishing collaborations which have been ongoing to the present day. The China Medical Informatics Association (CMIA) was founded in 1981 [18], helped organize MEDINFO 89, as did Singapore, which sponsored the China-Japan Medical Informatics Symposium in 1991, twelve China Medical Informatics Conferences since 1999, and a series of Pacific Health Informatization Summits since 2006 – and it is now hosting MEDINFO 2017 in Xiamen.

KOSMI, the Korean Society for Medical Informatics hosted MEDINFO 98 in Seoul, providing a major impetus for the development of biomedical and health informatics in the Asia-Pacific, which had by then formed its own regional association – APAMI, which held its inaugural conference in Singapore in 1994. Nursing Informatics (NI), which has its own Special Interest Group within IMIA, has flourished in the Asia-Pacific region, where KOSMI-NI hosted the 9th International Congress on Nursing Informatics in Seoul, from which the leadership developed a plan for the future of the field [19]. This is just one notable example of how coordination between IMIA, APAMI, the specialty groups, and national societies ensures the continuing vibrant growth of our interdisciplinary field of biomedical and health informatics throughout Asia, with strong international ties worldwide.

Results

IMIA 50th Anniversary Book Outline

The result of all the meetings organized or sponsored, in whole or part by the IMIA History WG over the past years, has been the gradual development plans for the publication of the IMIA History in 2017. As of this writing (April 2017), a large number of individual contributions have been already received, and a final set of invitations sent out to cover more comprehensively not only individual members, but also the histories of member societies and groups affiliated with IMIA: regional and national societies, special interest groups, and working groups. A schematic outline of the 50th Anniversary book is as follows:

PART 1: Collection of Short Personal Narrative or Individual Stories (about 2 to 5 pages in length). Each story is intended to address questions of how individual researchers and practitioners got started in biomedical and health informatics, and how their own work and engagement in the discipline unfold from their individual perspective: who inspired them in their work? what were the main ideas? What were challenges and informatics projects that got them started in the field? together with a brief reference to their main changes in direction since then. Summarized current reflections on the outcomes of early work from today's perspective were encouraged. With the above in mind, and at the request of many contributing authors, templates were developed to suggest questions that the personal stories could address, such as:

- Who inspired you to work in BMHI, or what circumstances determined your choice and how did this happen?
- What motivated you to get started and what were the main ideas or examples of biomedical science, information, and technology that inspired you?
- What were your initial expectations from your first project and what were the challenges you faced in realizing them?
- How did your first experiences in BMHI influence what you did next, and have done since? Summarize briefly your major changes in research, practice, or educational approaches, and their impact on your informatics contributions.
- What have been other experiences and professional position opportunities affecting your life work in BMHI?
- What are the social connections and relationships that were essential to your work, and how have these changed over time and with new technologies?
- How have the international dimensions of your work promoted or contributed to collaborations in research,

practice, education, or professional activities in organization and publications, conferences or other meetings, either locally or nationally? and did they have broader impact at regional and international levels for significant advances in BMHI?

- Other significant individual critical insights on BMHI developments beyond the above.
- From the perspective of 2016, what do you see as key shifts in “grand challenges” that you have dealt with in our discipline, and how have these inspired or challenged your work, and helped your personal and professional development?

Only a few citations to past publications have been requested, focusing on the most essential original and/or most recent summaries or reviews, so as to provide guideposts to any interested reader to pursue further searches online.

PART 2: Summaries of national and regional BMHI organizational short history entries, authored by country or region, from an informatics-related society or publication that authors have participated in – preferably in a leadership position, or with close participation and knowledge of the society.

A suggested template for Part 2 essays a requested by prospective authors suggest addressing the following:

- Briefly list major precursor work in underlying or related disciplines that influenced the emergence of BMH informatics, and briefly list major people influencing this.
- Which were the first contributors to the field in your country, with their institutions and fields of BMHI activity and major project titles (where known)?
- What were the major topics addressed by early informatics work in your country?
- When and how was informatics organized professionally in your country or region – relation to IMIA (and other national or international societies and institutions when relevant)?
- How did informatics evolved over its first 2-3 decades in your country/region – concentrating on the early days of the field, but bringing the short summary history to the present when possible.
- Short chronological listing of major early professional leaders in your country/region with roles, related to institutional affiliation and major projects.
- Brief listing of major conferences sponsored by professional BMH informatics societies in your country.
- Compare and contrast very briefly (a long paragraph at most) highlights of early vs. current work in BMHI in your country.

PART 3: Brief IMIA Working Group history entries, and short interpretive entries about publications and major research projects or programs of BMHI. The suggestions for this part of the book are included as a much shortened version of the above for professional societies.

Trends in MedInfo Contributions

A paper in the IMIA Yearbook of Medical Informatics, by one of the present authors [20], summarized the changes in full-paper contributions to MedInfo by broad subject areas and by first author geographic region up to 2004. We have reviewed the trends and how they have changed, by looking at similar statistical breakdowns for the 2007 and 2015 MedInfo, to see if there have been notable differences over the past decade.

The major subject areas were: 1) Health and Clinical Management, 2) Patient Records, 3) Health Information Systems, 4) Sensor, Signal & Imaging Informatics, 5) Decision Support, Knowledge Representation and

Management, 6) Educational and Consumer Informatics, and 7) Bioinformatics.

The total of full papers presented at Medinfo International Congresses seen to have grown steadily in the past decade compared to previous decades – for the Medinfo 2015 there were 512, for the 2007 there were 459, in contrast to 302 for the Medinfo 2004, and around mid or upper 200’s or low 300’s for the previous two decades.

The percentage of total for each of the categories is as follows:

1. Health & Clinical Management: 13.1% (2007) vs. 12.1% (2015)
2. Patient Records 18.3% vs. 14.8%
3. Health Information Systems: 13% vs. 23.2%
4. Sensors, Signals and Imaging Informatics: 4% vs. 3%
5. Knowledge Representation & Management, NLP, Decision Support: 17.6% vs. 19.3%
6. Education and Consumer Informatics: 21.6% vs. 19.1%
7. Bioinformatics: 1% vs. 1%.

We can see from the above list, that the only major changes over the 8 years were between Brisbane and Sao Paulo was that Patient Records (PR), which decreased somewhat, and Health Information Systems (HIS) increased by about the same amount. The latter is largely because the Data Bases category was included in HIS rather than PR, since the DB is more a systems function and can cover all kinds of databases, not just patient records. The jump in NLP is what accounts for the most of the rise in category 5 in 2015. Most of the other areas remain about at the same level.

The breakdown by the six broad subject areas indicated above, as computed by percentage of total papers, indicates certain trends that correlate with recent technological advances in mobile computing, interoperability, and databases, as well as the persistence of emphasis on computer decision support and knowledge representation (ontologies) and management, with a recent (MedInfo 2015) growth in the area of artificial intelligence, which specifically correlates with international trends in applications of machine learning for big data in all kinds of areas, including specifically clinical data. Notable also is the increase of papers contributed in the subarea of natural language processing (NLP), which reflects recent advances in this field.

In health informatics, there continue to be steady contributions in the areas of information systems for clinics and hospitals, and the field of nursing informatics has increased its participation in Medinfos very noticeably in the past couple of decades. The number of papers submitted on electronic patient records grew dramatically in 2007, by over 30% above the 2004 contributions, reflecting initiatives in this direction worldwide, and the especially strong impetus in the USA during this period; however, in 2015, had gone down again.

In education and consumer informatics, there has been a noticeable increase, including, not surprisingly, the new subarea of social media.

On the other hand, Medinfos have seen a continuing, but persistently low participation by authors submitting in the more scientific and technical subjects of sensor, signal and imaging informatics, and also bioinformatics, which reflects the difficulties of attracting these communities to contribute to general biomedical and health informatics, when most of their advanced discussion and research impact comes from presenting and publishing in their own specialty conferences.

The geographical distribution of contributions to Medinfo meetings over the years provides ample support for the early decision of IMIA to rotate the locations around the world, so

that all regions would have a chance to enjoy the hosting in succession, and encourage international participation most amply in the spirit of the Association. An overview of contribution breakdown, by percentage of papers in recent Medinfos, confirms the patterns of contributions that were noted earlier [20]. Looking at the 2007 Brisbane meeting, we see that the Asia-Pacific (AP) region was represented by 23% of contributors of full papers (upon which this comparison is made). This is approximately double the percentages of AP author contributions for Capetown 2010 (13%) and Sao Paulo 2015 (12%). European contributions, which have usually been the most substantial in number to all Medinfos outside those held in North America (NA), when NA attendance has predominated, showed the same for Brisbane (42% vs. 31% for NA) and Cape Town (at 51% vs. 24% for NA contributors), but not for Sao Paulo (29%), where North American authors contributed the most (33%). For the Sao Paulo Medinfo in 2015, Latin American contributions rose to 24%, in contrast to the regional contributions to the Congresses at other locations earlier, where their contributions have usually been between 1% and 6% (in Cape Town), demonstrating not only the rise in medical and health informatics in Latin America, but also the critical importance of rotating the locations. In a somewhat, but even more extreme parallel, the African first-author full paper participation at Cape Town, although still low at 7%, was seven times greater than participation at other Medinfo venues, when it never exceeded 1%, thus reflecting both the very restricted impact of health informatics in the continent, as well as the overwhelming organizational and financial difficulties of travelling for potential participants and practitioners from most of the continent. The Asia-Pacific participation, though reflecting a very different dynamic due to the advanced development of biomedical and health informatics in many of the countries, does, in all likelihood, reflect the factor of cost and convenience of travel to far-off venues, since participation in the Tokyo 1980 meeting was 35%, in Seoul 1998 was 25%, and Brisbane 2007 was 23%, all are predominantly higher than Asian-Pacific contributions to Medinfo meetings held elsewhere, such as Amsterdam 1983 (13%), Washington 1986 (7%), London 2001 (16%), San Francisco 2004 (7%), Cape Town 2010 (13%), and Sao Paulo 2015 (12%). We can therefore reasonably expect that the upcoming Medinfo 2017 in China will confirm this pattern and the importance of the regional rotations, as a wise policy of Medinfo organization for developing a world community of shared research, practice, and education, by encouraging the participation and dissemination of the most recent results in the field of biomedical and health informatics, dating back to its earliest days.

Discussion

For the IMIA History book progress, we can report that, from the contributions received so far, some authors have followed the template suggestions, and providing systematic accounts of the development of their professional lives and contributions to the field. Others have written very reflective and insightful (and sometimes even poetic) stories about the most important inspirations and aspects of their lives, and how they have affected their work in biomedical and health informatics. The present paper provides some preliminary perspective on the activities of the IMIA History WG, from which one can extrapolate for the future, when we expect to continue requesting contributions for further sets of papers on the history of our field for the IMIA Yearbook and other journal publications, for the History MediaWiki, and for potential further collected volumes on more specific topics of

research, practice, and education as experienced by those in our field.

Conclusion

The enthusiasm and progress in obtaining contributions from pioneers and leaders in biomedical and health informatics is a good sign that the 50th Anniversary IMIA History will become a unique original reference publication for future historians to study and derive their own conclusions about the very diverse motives, inspirations, and professional journeys of practitioners, technologists, and educational specialists in our field. The essays received so far have excellent original content, and eloquent writing, which reveals patterns of thought and argument that should be of great interest to present practitioners, as well as for future investigators and historians in our field. We look forward to completing this project as a first step in ongoing documentation, archiving, and historical writing about the development of international biomedical and health informatics. It will highlight how IMIA and its constituent societies have played a central role in spanning the global diversity of the field, encouraging scientific and professional collaborations, and providing a framework for international coordination and the development of the friendships that are so essential to this, as we approach the 50th Anniversary of the founding of its predecessor organization IFIP-TC4, to be celebrated in China, on the occasion of Medinfo 2017.

Acknowledgements

The authors gratefully acknowledge the encouragement and support of the IMIA Board and General Assembly and the contributions of their many friends and colleagues from around the world in the preparations for the 50th Anniversary IMIA History of International Biomedical and Health Informatics. They also wish to thank the reviewers for their many helpful suggestions.

References

- [1] H.E. Peterson and P. Lundin, Documenting the Use of Computers in Swedish Health Care up to 1980. *IMIA Yearbook of Medical Informatics*, Schattauer, Stuttgart, 2011: 169-174.
- [2] J.H. van Bommel, People and Ideas in Medical Informatics A Half-Century Review. *IMIA Yearbook Med Informatics*, Schattauer, Stuttgart, 2011:175-182.
- [3] M. Ball, Back to the Future: What Have We Failed to Learn? How Does the Future Look? *IMIA Yearbook Med Informatics* 2011: 196-198
- [4] V. Yacubsohn, A Short History of the Beginnings of Hospital Information Systems in Argentina, *IMIA Yearbook Med Informatics*, Schattauer, Stuttgart, 2012:163-168.
- [5] M. Fieschi, The Development of AI in Medicine and the Research Environment of the SPHINX Project at the Start of the 1980s, *IMIA Yearbook Med Informatics*, Schattauer, Stuttgart, 2012:153-160.
- [6] M.J. Ball, J.H. van Bommel, and S. Kaihara, IMIA Presidential Retrospectives on Medical Informatics, *IMIA Yearbook Med Informatics*, Schattauer, Stuttgart, 2007: 165-175.
- [7] D.A.B. Lindberg, The Evolution of Medical Informatics, in Myers JD (ed) *Medical Education in the Information Age .Proc SCAMC*, Assoc Amer Med Colleges, 1986: 86-95
- [8] C.A. Kulikowski, An Opening chapter of the first generation of Artificial Intelligence in Medicine: The First Rutgers AIM Workshop, June 1975, in *IMIA Yearbook of Medical Informatics 2015* (Jaulent MC, Lehmann C, Serrousi B, eds), IMIA and Schattauer GmbH, Stuttgart, 2015: 227-233.
- [9] G.I. Mihalas, *An attempt to define major stages of Medical Informatics development*, CD Proceedings IMIA-LAC Conference, Guadalajara, Mexico, 2011.
- [10] C. Kulikowski and C. McGrew, An Informatics Approach to Chronically the History of IMIA. *Proc Medinfo* 2013: 1130.
- [11] G.I. Mihalas, J.H. van Bommel, B. Richards, R. Engelbrecht, I. Masic, R. Haux, and C. Kulikowski, *Medical Informatics History - A View from*

- Socio-Political Context in European Countries*. Workshop MIE, Istanbul, 2014.
- [11] G.I. Mihalas, J. Zvarova, and C. Kulikowski, *European History of Medical Informatics*, Proc. EFMI STC, Prague, 2013. Available at: www.stc2013.org/joint-activities/panels
 - [12] G.I. Mihalas, J. Zvarova, C. Kulikowski, M. Ball, J.H. van Bommel, A. Hasman, I. Masic, D. Whitehouse, and B. Barber, History of Medical Informatics in Europe - a Short Review by Different Approach. *Acta Inform Med* **22**(1) (2014), 6-10.
 - [13] G.I. Mihalas, Evolution of Trends in European Medical Informatics. *Acta Inform Med* **22**(1)(2014), 37-43.
 - [14] G. Deželić, J. Kern, M. Petrovečki, V. Ilakovac, and M. Hercigonja-Szekeres, Medical Informatics in Croatia- a Historical Survey. *Acta Inform Med* **22**(2014), 49-59.
 - [15] I. Masic and G. Mihalas, *Contributions to the History of Medical Informatics*, Avicenna, Sarajevo, 2014.
 - [16] A. McCray, ACMI at 30 – The Founding of the American College of Medical Informatics and the American Medical Informatics Association, *Proc. AMIA 2014*, Washington, DC, 2014.
 - [17] D.A.B. Lindberg and S. Kaihara (Eds), *MEDINFO 80 – Proceedings of the Third World Congress on Medical Informatics*, North-Holland, Amsterdam, 1980.
 - [18] J. vanBommel and A. McCray (Eds), *IMIA Yearbook of Medical Informatics 93*, Schattauer, Stuttgart, 1993.
 - [19] P.J. Murray, H.A. Park, W.S. Erdley, and J. Kim (Eds), *Nursing Informatics 2020: Towards Defining our Own Future*, IOS Press, 2007.
 - [20] C.A. Kulikowski, IMIA: *Coalescing Medical Informatics Worldwide for 40 Years*, IMIA Yearbook Med Informatics, Schattauer, Stuttgart, 2007: 176-185.

Address for correspondence

Casimir A. Kulikowski. Department of Computer Science
Hill Center, Busch Campus, Rutgers University, Piscataway, NJ
08855, USA. Email: kulikows@cs.rutgers.edu (preferred)
Phone: 732-277-3803; Fax: 732-445-0537

Towards a Usability and Error “Safety Net”: A Multi-Phased Multi-Method Approach to Ensuring System Usability and Safety

Andre Kushniruk^{a,b}, Yalini Senathirajah^c, Elizabeth Borycki^a

^a School of Health Information Science, University of Victoria, Canada,

^b Department of Development and Planning, Aalborg University, Aalborg, Denmark

^c State University of New York, Brooklyn, USA

Abstract

The usability and safety of health information systems have become major issues in the design and implementation of useful healthcare IT. In this paper we describe a multi-phased multi-method approach to integrating usability engineering methods into system testing to ensure both usability and safety of healthcare IT upon widespread deployment. The approach involves usability testing followed by clinical simulation (conducted in-situ) and “near-live” recording of user interactions with systems. At key stages in this process, usability problems are identified and rectified forming a usability and technology-induced error “safety net” that catches different types of usability and safety problems prior to releasing systems widely in healthcare settings.

Keywords:

User-Computer Interface; Patient Safety, Health Information Systems

Introduction

Ensuring the usability and safety of healthcare IT is essential in order to lead to successful adoption of healthcare IT. In recent years, the issue of poor usability of many healthcare IT systems has come to the fore [1,2]. Indeed, in healthcare there are continued reports of end user dissatisfaction with systems and reports of serious usability problems. In addition, implementation failures due to usability issues and negative impacts of some systems on clinical workflow continue to be reported [3]. In recent years, poor usability has also been closely linked with the occurrence of technology-induced error. Technology-induced errors have begun to be studied more widely in health informatics and represent a class of errors that emerge when IT is deployed and used in complex healthcare settings [4]. Such errors are difficult to find using traditional software testing methods (e.g. white- and black-box testing) and in many cases are only detectable (given conventional testing methods) when healthcare IT is operational in real healthcare contexts [4].

Many different approaches have been proposed and applied for testing the usability of systems and for evaluating their impact on patient safety. Each method has its advantages and disadvantages. For example, basic usability testing can be used to detect surface level usability problems with healthcare IT but may not be sufficient to assess how a system affects clinical workflow (e.g. real patient encounters) or complex clinical decision making and reasoning. In contrast, methods such as clinical simulation may allow for analysis of the

impact of systems on clinical work activities but may not detect basic usability problems. Therefore, a multi-method approach may be needed to ensure system usability and safety prior to release. Furthermore, integrating different methods in a multi-phase approach holds promise for obtaining the full benefits of the different methods. In this paper we will describe a methodological framework for developing a “safety net” approach to detecting and rectifying usability problems and potential technology-induced error prior to widespread release of healthcare IT. It will be argued that the methods described in this paper need to be considered not only in the context of system design (e.g. during one-time design of vendor-based systems), but also in subsequent implementation of commercial systems (e.g. implementation of vendor products, such as electronic health records (EHRs) in hospital settings). As will be described, the methodological approach outlined in this paper has been successfully applied to both improving the design of healthcare IT developed by developers and vendors, as well as assessing the usability and safety of completed healthcare IT deployed in healthcare organizations.

Methods

A process for testing healthcare IT was developed by the authors and involves several phases described below. In the first phase, rapid low-cost usability testing is conducted. Video recording is used to collect data on typical user interactions, where users are recorded while carrying out representative tasks using a system under study [5]. Analysis of the resultant video data is facilitated by using principled coding schemes and computer coding tools. This is followed by a phase of testing under simulated conditions (i.e. by applying clinical simulation methods). Where possible, this is then followed by testing of live, or near-live clinical interactions, where users are observed under close-to-real or real conditions and settings [6]. After fixing errors and problems detected at each of these stages, the system is eventually released for widespread deployment.

The authors have employed the approach described above in the design and development phases of the system life cycle as well as for testing and modifying vendor-based systems that were purchased by healthcare organizations prior to their widespread deployment. The intent is to form in essence a usability and error detection “safety net” that can catch different types of usability problems and errors before a system is released on a widespread basis. Details of our methodological approach are described below.

Phase 1. Low-Cost Rapid Usability Testing

The detection of major surface level usability problems is key to ensuring overall system usability and safety, and although not sufficient on its own to ensure system safety, it is a critical first stage. An approach we have developed and applied is termed "low-cost rapid usability engineering" [5]. Using this approach, free or low-cost screen recording software is installed on computers used in the testing. Participants (e.g. health professionals, patients etc.) are asked to carry out representative tasks using the system under study. The interactions of the study participants are recorded (i.e. screen recording, audio recording and video recording of physical actions) in their entirety. Participants may also be asked to "think aloud" or verbalize their thoughts while interacting with the system. Approaches may include collecting user interaction data (e.g. computer screen interactions as movie files) along with audio of participant verbalizations. This approach to usability testing has been used in many studies and in our approach constitutes the first phase of the "safety net."

Phase 2. Error Correction from Usability Testing

The data collected in Phase 1 consists of digital screen recordings of one or more users (i.e. study participants) interacting with a system under study while verbalizing their thoughts [7]. One approach to accelerating the process of analyzing this data has been to use predefined coding schemes, such as the one developed by the authors that provides categories of commonly detected usability issues [8]. Such schemes (in conjunction with video coding tools) allow for replay of the video data and coding of that data for key usability problems in close to real-time [5]. Identified usability problems are prioritized and presented to the development or implementation team in the case of vendor-based systems being deployed in healthcare settings. Decisions regarding which issues or problems to rectify at this point are discussed with the IT staff, taking into account the severity of the problems as well as other contingencies such as the release schedule and resources for modifying the system. In a number of our studies, we have found this stage to have a quick turnaround time, but system modifications may need to be restricted to fixing problems considered most urgent or severe. At this stage, issues identified are typically surface level usability problems.

Phase 3. Clinical Simulation

From our early experiences, it was found that Phase 2 usability testing alone is not enough to ensure the usability and safety of healthcare IT. As a consequence, the authors and colleagues have worked on developing low-cost rapidly-conducted clinical simulations [4]. We have defined clinical simulations as an extension of usability testing that includes not only observing representative users doing representative tasks, but also testing in representative or real environments (e.g. clinics with real or realistic workflow) [9-12]. This may involve using real (or realistic) testing scenarios in testing. Clinical simulations in our work have varied from use of actors to development of digital video patients used in testing interactions involving real or highly realistic tasks and workflows. For example, a clinical simulation may involve participants (e.g. physicians) interviewing a simulated patient while using a new electronic health record system or decision support tool to enter and record patient data. Data collection involves video and audio recording of all user interactions with systems under study.

Phase 4. Error Correction from Clinical Simulation

The analysis of the data collected from clinical simulations (e.g. video and audio data) is similar to that of the analysis of usability testing data. However, at this phase, the focus of analysis is typically on assessing the impact of systems on clinical workflow and healthcare activities. Coding schemes we have employed at this phase include categories for assessing impact of systems on decision making, reasoning, physician question asking, and also categories for identifying potentially negative impact such as increased time to complete tasks and workarounds that are potentially dangerous. In addition, from our experience we have found that surface level usability problems not detected from Phase 1 testing may also end up being identified in this stage. Major defects (e.g. negative system impact on clinical workflow) identified from clinical simulation are corrected, where feasible, before moving on to the next phase.

Phase 5. Near-Live Testing

In a number of studies we have conducted, after clinical simulation has been completed, we have moved to observe and record a limited number of user interactions with the system in a live setting, prior to widespread deployment or release [6]. This limited naturalistic testing involves setting up unobtrusive recording equipment (e.g. using free screen recording software installed on systems used) and allows us to collect data on use of the system in real contexts prior to full-scale product release or healthcare organization implementation. Issues at this stage include obtaining institutional review board consent for collecting data that may involve real patients. Whether this additional level of testing is feasible depends on gaining organizational access to settings and contexts in which a system will be ultimately deployed (e.g. a hospital ward, or even an operating room off hours).

Phase 6. Error Correction from Near-Live Testing

From our work we have found that limited naturalistic testing almost invariably uncovers further issues that could not have been found from usability testing or clinical simulation alone. Major concerns or issues that are identified are brought to the attention of the design/implementation teams, with problems warranting immediate fixing being rectified prior to widespread system release.

Phase 7. Widespread Deployment and Continual Monitoring

Upon completion of the phases described above, the system can be released on a wide scale with greater confidence that major issues have been detected before "going live". This may involve releasing a healthcare IT software product to the market, or the widespread deployment of a vendor system throughout a hospital or health region. Continual monitoring of system use, in terms of possible usability problems or errors, is recommended. As shown in Figure 1, the safety net approach can be considered a cyclic process, with each phase leading to the correction of different types of errors. Figure 1 also shows the need for continual error monitoring (and correction), once the system is in use.

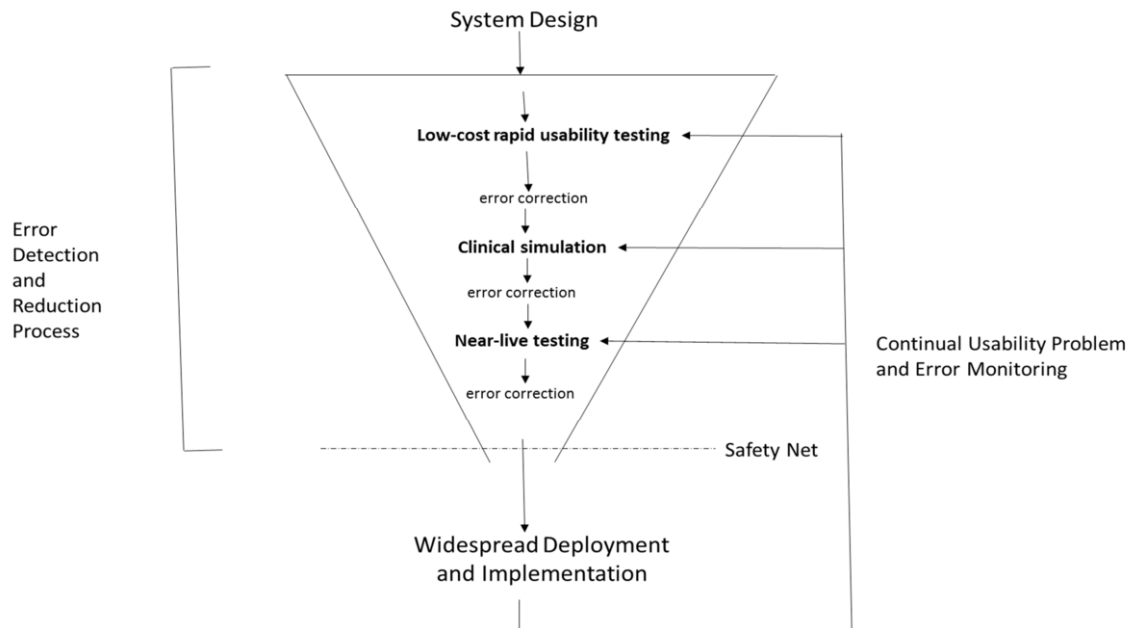


Figure 1—A safety net approach for usability problem and error detection.

In Figure 1 we can also see that the approach attempts to identify and rectify as many usability problems and errors as possible before widespread deployment and implementation. At the top of Figure 1, we can see that after initial system design is complete, low-cost rapid usability testing can be employed to correct surface level usability problems. This is followed by clinical simulation and near-live testing. At each stage different types of errors are identified, with surface level usability problems typically detected during low-cost rapid usability, while issues around workflow are more likely to be detected in subsequent phases. The dotted lines in Figure 1 illustrate the “sieve” part of the safety net, which is designed to catch as many errors as possible before a system goes to widespread deployment, as illustrated in the lower part of Figure 1. Once deployed, continual monitoring for usability and feedback from error reports from end-users may need to iteratively trigger further cycles of usability analyses, as indicated by the arrows in the right side of Figure 1.

Results

We have applied the multi-method, multi-phased approach described in this paper in a number of assessments of EHRs, decision support systems (DSSs) and mobile health applications. A finding emerging from our work is that the types of errors and usability problems detected at each phase are different, and that therefore, a multi-phased approach was needed to detect the widest possible range of problems. For example, early phase usability testing (in Phase 1) typically uncovered surface level usability problems, while testing

involving clinical simulations (Phase 2) typically detected issues related to impact on clinical workflow, including identifying potentially dangerous workarounds. In Phase 3, near-live testing typically revealed a range of contextual issues that would not have otherwise been predicted by the previous phases. It should also be noted that surface level usability problems (and even software programming errors not caught by traditional white- and black-box testing) were typically detected throughout all phases of the approach (i.e. even during clinical simulations and live or near-live testing, usability problems would still be detected that had not been found in Phase 1). The approach thus provided a usability and technology-induced error “safety net”, without which serious usability problems and errors would have found their way into actual system use and patient care.

In development and refinement of healthcare IT during the design and prototyping phases, our early work employed components of the approach described in this paper, specifically classic usability testing of an EHR was followed by clinical simulation of a commercial EHR system, where actors played the part of a patient while study participants (e.g. physicians) interacted with the system to handle patient cases [13].

In more recent work, we are adapting some of the approaches described in this paper to the testing of a novel, composable user interface designed to allow for near-live testing using remote usability engineering methods [14]. This will take place in two phases, the first being detailed, comparative, in-person, near-live testing in crossover studies of commercial and experimental systems, in which the user is asked to create interfaces in the composable system, and both systems are

evaluated as to user behavior and diagnostic accuracy. Another phase is testing larger numbers of subjects via remote near-live testing, in which software working in the background records user actions and generates a video. The ability to send users a link and have them perform tests at a distance, without in-person meetings, allows for larger numbers of subjects and testing of sharing functionality, which is expected to yield new insights about composable systems and clinician cognitive processes and communication.

The approach described in this paper has also been applied to ensure that the implementation of vendor-based products (e.g. EHRs) as well as their extensions (e.g. addition of decision support capabilities to an EHR) do not inadvertently introduce serious usability or safety issues when implemented in hospitals. Along these lines, we have successfully employed parts of the approach as well as the full approach on a number of projects. In one study, we employed the full multi-phased approach in the implementation of clinical guidelines within a commercial EHR product [6]. In Phase 1 of the approach described in this paper, basic usability testing was conducted. This testing identified a range of surface level usability problems, such as navigational issues, consistency problems and the need for improved layout and wording of alerts. This included the need to relabel alerts that appeared on the screen and to make them more conspicuous to the physician's eye (these issues were rectified in Phase 2). In Phase 3, a clinical simulation was conducted involving participants (8 physicians) interviewing a simulated digital patient (from which they could ask for patient information) in order to assess the impact of the system on clinical workflow. Analysis of the video data from these simulations (conducted in Phase 4) indicated the introduction of the guidelines inadvertently changed physician workflow in a number of unexpected ways. For example, some participants found that the alerts were triggered too early, before the participant had a chance to review the patient's information and data. As a result, alert triggering mechanisms were modified during Phase 4. In Phase 5, limited naturalistic testing was conducted and defects detected from this phase were rectified in Phase 6, prior to widespread release of the guidelines. A subsequent clinical trial showed a high rate of uptake of the guidelines by end-users in the hospital, in large part due to the multi-phased multi-method testing that was conducted prior to system release.

In another application of the approach, we employed these methods in the assessment of a medication administration system in a Japanese hospital. Phase 1 testing involved having 16 physicians and nurses interact with the system to carry out medication tasks while being video recorded. Phase 1 identified a number of specific usability problems, while issues with workflow as a result of the system (as well as potential technology-induced errors) were identified in Phase 3 using clinical simulations. In terms of problems at the workflow level, it was found that the system was inflexible and rigid during emergency situations, when there was a need to override the systems due to time constraints. Results of the phased multi-method approach led to refinement of the system to include emergency overrides, improved workflow and streamlined usability during the defect correction phases of the process [15].

Discussion

In our work we have integrated a number of methods for ensuring system usability and safety. We found that each phase of this method detected different types of problems,

starting with surface level usability problems and leading to issues related to workflow, impact on physician cognitive processes, and patient safety. It is argued that such a multi-method phased approach is needed to create a "safety net" to detect and rectify a range of errors and problems prior to widespread system release. The approach described in this paper is beginning to be applied in healthcare organizations where usability and safety have been deemed of utmost importance in system implementations and deployments. Current areas we are working on include streamlining the methods and processes involved to shorten the time taken for analysis (using published coding schemes and computer tools) and increasing dissemination of the approach through demonstration projects at key healthcare organizations worldwide.

Conclusions

The usability of healthcare IT continues to be a problem, with reports of systems that are unusable and in some cases unsafe. One approach has been the development of guidelines and regulations to improve the situation [17-19]. In another direction, to address these issues, the authors have developed a methodological approach that involves several phases of data collection and analyses. This approach begins with low-cost rapid usability testing, followed by a phase of clinical simulation and finally near-live or live testing involving a limited number of participants. Our experiences with this approach has indicated that by applying this "safety net" approach, a wide range of issues and problems can be detected (and rectified) prior to widespread release of healthcare IT systems. Areas of future work include demonstrating the cost-effectiveness of the approach and disseminating information about it more widely. This method can become incorporated in healthcare IT practice as there is a need to enhance patient safety through improving the usability of healthcare IT.

Acknowledgements

This work is supported by the Agency for Healthcare Research and Quality (AHRQ) grant R01HS023708.

References

- [1] B. Middleton, M. Bloomrosen, M.A. Dente, B. Hashmat, R. Koppel, J.M. Overhage, T. H. Payne, S.T. Rosenbloom, C. Weaver, J. Zhang, Enhancing patient safety and quality of care by improving the usability of electronic health record systems, *J Am Med Inform Assoc* **20** (2013), e2-8.
- [2] A.L. Kellermann, S.S. Jones, What it will take to achieve the as-yet-unfulfilled promises of health information technology, *Health Affairs* **32**(1) (2013), 63-68.
- [3] E. Jamoom, P. Beatty, A. Bercovitz, et al. Physician adoption of electronic health record systems: United States, NCHS data brief, no. 98. Hyattsville, MD: National Center for Health Statistics, 2012.
- [4] E. Borycki, A. Kushniruk, C. Nohr, H. Takeda, S. Kuwata, C. Carvalho, M. Bainbridge, J. Kannry, Usability methods for ensuring health information technology safety, *Yearb Med Inform* **8**(1) (2013), 20-27.
- [5] A.W. Kushniruk, E.M. Borycki, Low-cost rapid usability engineering: designing and customizing usable healthcare information systems, *Health Q* **9**(4) (2006), 98-100.
- [6] A.C. Li, J.L. Kannry, A. Kushniruk, D. Chrimes, T.G. McGinn, D. Edoynabo, D.M. Mann, Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support, *Int J Med Inform* **81**(11) (2012), 761-72.
- [7] A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* **37**(1) (2004), 56-76.

- [8] A. Kushniruk, E. Borycki, Development of a video coding scheme for analyzing the usability and usefulness of health information systems. *Stud Health Technol Inform* **218** (2015), 68-73.
- [9] E. Borycki, E. Keay, Methods to assess the safety of health information systems, *Healthc Q* **13** (2010), 47-52.
- [10] A.W. Kushniruk, C. Nohr, S. Jensen, E.M. Borycki, From usability testing to clinical simulations: Bringing context into the design and evaluation of usable and safe health information technologies, Contribution of the IMIA Human Factors Engineering for Healthcare Informatics Working Group, *Yearb Med Inform*, 2012 Dec;8:78-85.
- [11] S. Jensen, S.L. Rasmussen, K.M. Lyng, Use of clinical simulation for assessment in EHR-procurement: design of method, *Stud Health Technol Inform* **192** (2013), 576-80.
- [12] A.W. Kushniruk, E.M. Borycki, S. Kuwata, J. Kannry, Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments, *Stud Health Technol Inform* **169** (2010), 915-919.
- [13] V.L. Patel, A.W. Kushniruk, S. Yang, S., J.F. Yale, Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning, *J Am Med Inform Assoc* **7**(6) (2000), 569-585.
- [14] B. Albert, T. Tullis, D. Tedesco, Beyond the usability lab: Conducting large-scale online user experience studies, Morgan Kaufman, New York, 2010.
- [15] A.W. Kushniruk, E.M. Borycki, S. Kuwata, J. Kannry, Predicting changes in workflow resulting from healthcare information systems, *Healthc Q* **9** (2006), Oct 15:9(Sp).
- [16] National Institute of Standards and Technology, Health IT Usability, Accessed (November 8, 2016) from: <http://www.nist.gov/healthcare/usability/>
- [17] Agency for Healthcare Research and Quality, EHR Usability Toolkit. Accessed (November 8, 2016) from: https://healthit.ahrq.gov/sites/default/files/docs/citation/EHR_Usability_Toolkit_Background_Report.pdf
- [18] Office of the National Coordinator. Safer guides will help optimize safety. 2014 Accessed (November 8, 2016) <https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/safer-guides-optimize-safety/>
- [19] D. F. Sittig, J.S. Ash, H. Singh, ONC issues guides for SAFER EHRs. *J AHIMA* **85**(4) (2014), 50-52.

Address for correspondence

Corresponding author: Dr. Andre Kushniruk, Professor and Director,
School of Health Information Science, University of Victoria,
Victoria, Canada, email: andrek@uvic.ca

The New HIT: Human Health Information Technology

Tiffany I. Leung, MD, MPH^a, Mary K. Goldstein, MD, MS^{b,c}, Mark A. Musen, MD, PhD^d, Ruth Cronkite, PhD^{b,e}, Jonathan H. Chen, MD, PhD^f, Assaf Gottlieb, PhD^g, Eran Leitersdorf, MD^{h,i}

^a Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

^b Department of Veterans Affairs, VA Palo Alto Health Care System, Palo Alto, CA, USA

^c Center for Primary Care and Outcomes Research, Stanford University, Stanford, CA, USA

^d Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA, USA

^e Department of Sociology, Stanford University, Stanford, CA, USA

^f Division of General Medical Disciplines, Stanford University, Stanford, CA, USA

^g School of Biomedical Informatics, University of Texas Health Science Center, Houston, TX, USA

^h Center for Research, Prevention, and Treatment of Atherosclerosis Internal Medicine Department, Hadassah Hebrew University Medical Center, Jerusalem, Israel

ⁱ Clinical Informatics, Stanford Health Care, Palo Alto, CA

Abstract

Humanism in medicine is defined as health care providers' attitudes and actions that demonstrate respect for patients' values and concerns in relation to their social, psychological and spiritual life domains. Specifically, humanistic clinical medicine involves showing respect for the patient, building a personal connection, and eliciting and addressing a patient's emotional response to illness. Health information technology (IT) often interferes with humanistic clinical practice, potentially disabling these core aspects of the therapeutic patient-physician relationship. Health IT has evolved rapidly in recent years – and the imperative to maintain humanism in practice has never been greater. In this vision paper, we aim to discuss why preserving humanism is imperative in the design and implementation of health IT systems.

Keywords:

Humanism; Electronic Health Records; User-Computer Interface

Introduction

“Good morning Mr. Jones, it’s great to see you today for your check-up. Tell me, what can I do for you today?” While Dr. Smith is logging in to the in-room workstation, Mr. Jones, who has diabetes and hypertension, begins describing his flu-like symptoms. He is interrupted by a text message from his spouse reminding him to discuss concerns about sexual function during the visit also. After describing his current symptoms, he also is traveling to Florida during the winter months where he has a second home, and plans to travel to Brazil, a lifelong item on his bucket list. He has questions about travel immunizations, and would like to talk about how concerned he should be about Zika virus, which he has heard about on the news for months.

In the meantime, Dr. Smith notes multiple high-importance alerts reminding her to address blood pressure higher than the normal range today, and that Mr. Jones is overdue for his first pneumococcal vaccine, monitoring labs and an eye exam for diabetes. Additionally, his chart still shows he is due for colon cancer screening, which he had previously declined. She also notes high-priority indicators about out-of-range lab values for other patients that just arrived in her in-basket, and receives a

repeat page to her smartphone’s secure text messaging app from a radiologist about a critical read from an earlier appointment.

Background

It is no surprise that practicing clinicians are pressured by time constraints and high performance expectations when providing routine patient care. The average practicing primary care physician in the U.S. is limited to spending 15 minutes with each patient [1], yet they and their team may be expected to spend approximately 7.4 hours per working day addressing only preventive services for patients [2] and 3.5 hours per day to provide care for chronic conditions [3], spend on average 2.3 minutes responding to patient questions submitted via electronic messaging [4], perform other in-between visit care management and coordination tasks, and still be required to achieve high patient satisfaction scores. Adding burdensome documentation requirements for billing, quality measurement, and other secondary purposes not directly related to the provision of health care [5;6], and with difficult-to-use health information technology (IT) and electronic health records (EHRs) [7;8], only adds administrative burdens to the already time-consuming responsibilities of care.

Desktop medicine, consisting of activities related to care provision, such as progress note writing, documenting telephone encounters, addressing secure messages and prescription refills, reviewing results and placing orders, is estimated to occupy approximately half of a physician’s time in clinic [9]. The other half of clinic time is face-to-face time in office visits, as estimated by examining physician access logs for a large practice of internists, family medicine physicians, and pediatricians. A recent time-and-motion study of physician practices in multiple specialties was even more concerning, finding that physicians spent 27.0% of time on face-to-face time with patients and 49.2% of time on EHR and clerical work [10]. In another study, estimated face-to-face time may be more optimistic, that is, for every 3 minutes of face-to-face time with a patient, 1 minute is spent on clerical tasks, predominantly charting [11]. In the same study, in the exam room with patients, physicians spent 52.9% of that time on face-to-face time and 37.0% on EHR and clerical work. In surveys of physicians who use EHRs, more than one-third of physicians agree that EHRs decrease the quality of face-to-face patient interactions [12;13].

Beyond quantifying and describing in detail the administrative burdens of clinical practice, a call to return to the primary goal of clinical practice must drive the redesign of health information technology and its integration to once again support patient care. Healing and preserving patients' dignity, core tenets of the sacred Hippocratic Oath, guide the humanistic service of patient care. Over time, as these studies have observed and reported, the human elements of clinical medicine, these sacred oaths, physicians' ability to practice, and ultimately patients, are endangered by health IT designed in service to the secondary purposes – rather than the primary purpose – of care. In this vision paper, the authors summarize their perspectives on this tension between humanism and health information technology.

During two-year sabbatical for E.L., sponsored by Stanford Health Care's Chief Executive Officer, he established a monthly Clinical Informatics Discussion Forum for Year 1, which included up to 22 faculty and trainees from Stanford University and the Department of Veterans Affairs from September 2013 to September 2015. Forum members included practicing physicians, physician executives in clinical informatics, biomedical and clinical informatics researchers and educators, experts in clinical decision support design, a sociologist, and clinical informatics trainees. Each participant had titles for multiple roles, for example, E.L. is a practicing internist, educator, and clinical and biomedical informatics researcher. Not all participants were able to attend all meetings during Year 1. For some forum meetings, an invited guest from another discipline at Stanford University, including aeronautics, or affective (cognitive) psychology, would also participate in discussion. During the discussion forum, participants drew upon personal knowledge and expertise and identified known sources of literature in order to address key themes raised. Further focused literature searches were performed in a problem-oriented approach. In Year 2, the authors, all participants from the original discussion forum convened in Year 1, formed a subgroup that continued quarterly discussions that culminated in the writing of this vision paper. This vision paper reflects themes derived from the discussion forum and subsequent discussions by the authors.

Humanism in Medicine

Health information technology and electronic health records are accelerating the digitization of health care and challenging the practice of medicine to evolve. Vinod Khosla claimed in 2012 that technology "will replace 80 percent of what doctors do" [14]. The transition has only just begun, and growing pains, particularly the painful loss of the human touch or humanism in digital medicine, are rippling through clinical practice. Numerous reports describe disillusionment with wasted time and effort and growing discontent associated with computer-based documentation and clerical tasks. Health information technology, especially EHRs, directly interfere with – rather than enable – patient-physician communication and workflow during visits [7]. Moreover, excessive focus on rapidly developing and implementing IT innovations beyond EHRs may further erode humanism in clinical practice, especially when such innovations are inadequately evaluated. Correspondingly, the role of the physician in technology-enabled modern medicine must evolve to match changing care settings. To meet these needs, the universal guiding principle in designing any health IT system must be humanism. That is, we need human health information technology systems, which are centered on the basic human connection between doctor and patient and less on the novelty of technology alone.

Humanism in medicine is defined as health care providers' attitudes and actions that demonstrate respect for patients' values and concerns in relation to their social, psychological and spiritual life domains [15]. Specifically, humanistic clinical medicine involves showing respect for the patient, building a personal connection, and eliciting and addressing a patient's emotional response to illness [16]. However, health IT has enabled more interruptions and external requirements on individual encounters, which in turn threaten the quality of the physician-patient connection [17]. To preserve the physician-patient relationship, we as a medical community must prioritize humanism as our central tenet for driving the development of IT that fully supports and enhances – rather than disrupts – the patient-physician relationship. Additionally, the practicing physician must actively preserve the physician-patient relationship, attending primarily to key features of a patient's life experience without distraction from health IT and also through proficient IT system navigation. Thus, we must act now so that health IT sufficiently evolves into human health IT, intelligently supporting physician's role in patient care. This evolution towards human health IT is necessary to advance both to scientific and technological innovations while necessarily preserving humanism in clinical practice.

Why humanism matters

Social, psychological, and behavioral determinants of health are central to psychiatrist George Engel's biopsychosocial model of medicine described nearly 50 years ago [18]. These principles remain relevant to clinical practice and medical education today, as precision medicine necessarily incorporates individualized care based upon psychosocial characteristics of each patient [19].

Social and behavioral determinants of health are well-established determinants of patient outcomes [20]. Assessing each individual's health context is necessary to fully understand a patient's condition and personalize care recommendations. Currently, social determinants and a comprehensive context of health are not yet systematically collected through health IT, although there is increasing attention to developing frameworks to guide such work [20]. Even more streamlined surveys of health context, for example, the Adverse Childhood Experiences questionnaire, have been suggested as a tool to assist in identifying patients at highest risk for developing certain mental and physical health diagnoses [21], yet largely still remains to be systematically integrated into practice. As a consequence, traditionally, humanistic practices can help elicit this information, and for now should remain a high-priority skill for any practitioner to ensure these relevant features are incorporated into patient-centered care decisions.

Humanistic practice strengthens the continuous physician-physician relationship and facilitates identification of relevant patient features for care. Humanistic practices build patient confidence in the relationship and empower the patient as central stakeholder in the process of care. In principle, continuity settings allow physicians to attend to human dimensions of care, spend less time re-collecting potentially complex histories [22] and more time on addressing the present condition(s), interpret information within the individual's context of care, and engage in optimal decision making. Such settings facilitate the physician's motivation to invest the time [22] to build rapport with each patient by actively listening, eliciting patient concerns, and applying excellent communication skills, thus engaging patients as active participants in their health care [15]. While machine learning systems are evolving to identify patient features for decision-making, and data-driven clinical decision support to optimize

decisions, these still remain a goal for the future rather than a reality in real-time shared decision making.

Humanistic practice promotes optimal shared decision-making. An in-depth understanding of a patient's goals, values, and beliefs about their health is essential to informed and shared decision-making. Potentially complex decisions are increasingly common in the primary care setting. For example, more than two-thirds of Medicare beneficiaries in the US have multiple chronic conditions; for these patients, thoughtful attention and time to appropriately consider each condition and resolve conflicting treatment recommendations may be necessary. Strong interpersonal skills are also crucial when discussing goals of care or breaking bad news, such as informing a patient of a new or grave diagnosis [23]; these procedures require careful discussion of appropriate treatment considerations that are both evidence-based yet tailored to the preferences and goals of each patient. An established and trusting patient-physician partnership provides a solid foundation for facilitating difficult conversations and patient-centric decision making. Currently, a health IT system might provide access to useful information and enable such conversations but cannot alone provide such nuanced human dialogue without physician intervention.

How technology compromises humanism

EHRs are ubiquitous, with 78% of practices having implemented EHR by 2013 [24]. Unfortunately, current systems still interfere directly with practice by shifting finite yet vital cognitive resources – physician attention, memory, and time – away from the patient and towards the EHR. In the US, where nearly 50% of visits are performed by family practitioners, general internists, and pediatricians, and in-person visits are often limited to 15-minutes [1], reducing physician attention during an already constrained visit is unacceptable. Evaluation of the unintended consequences of health IT on humanism in practice and medical education is a potential safeguard in support of high-quality health care systems [25;26], including patient safety issues, medical errors, and disruption in clinical workflow, but health IT system implementation and the abundance of new and perhaps proprietary IT innovations outpace such evaluations. Consequently, humanistic practice and care quality are at risk of further erosion.

Technologizing the science of medicine alone threatens humanism. Biomedical knowledge is increasingly able to be represented in a computer system or health IT solution, but humanism cannot. While the technological approach assumes that a biomedical model of disease alone allows for a fully digital representation of a patient, it does not incorporate social determinants of health, patient preferences, or shared decision making. These highly relevant features of a patient and their health context are essential for understanding an individual's health and life context and for making informed and appropriate health care decisions.

Health IT distracts physicians from practicing humanistic care. Maintaining humanism in clinical practice requires focused time interacting with the patient. Health IT interferes with clinical workflow via inappropriate alerting and time-consuming data entry tasks, often directly contributing to cognitive overload, distraction, and reduced attention to important clinical information. For example, interruptions or multitasking frequently occur as a consequence of health IT implementation. Interruptions require additional cognitive resources and effort to reacquire and then subsequently perform prior tasks, potentially leading to an overreliance on intuitive or heuristic thinking and increasing the risk of systematic and unconscious biases in care. Distractions interfere with rational

thinking and clinical judgment when high degrees of uncertainty or many pieces of information require careful synthesis, as is often the cases in clinical encounters, but also for simple tasks such as patient education and counseling.

Health IT excludes humanism from potentially complex shared decision making. Patient preferences, goals of care, and social determinants of health provide a context for making difficult medical decisions. Given the cognitive resources and time to do so, physicians can engage patients in dialogue to elicit their care preferences and provide the necessary clinical context to allow for shared decision making. As noted previously, the digital patient may not sufficiently represent true patient preferences in decision making processes.

Recommendations

We propose guiding principles for the new practice of digital medicine and the design of health IT, towards human health IT, which prioritizes the physician-patient connection above all else.

1. *Physicians must continue to practice with a humanistic approach, with empathy and compassion towards the patient's condition and care context.* Human health IT also should support and enhance – and never interfere with – the patient-physician connection. Perturbations of thought, affect, time, and other valuable cognitive resources must be minimized to enable patients and physicians to focus on key biopsychosocial aspects of care. This principle is particularly high-priority in situations with a great deal of uncertainty or numerous complex issues to consider in decision making. In particular, patients with multiple chronic conditions or elderly patients, may have particularly inherent complex decision-making required for their care [27;28].
2. *Physicians must adapt to technology-enabled clinical practice by prioritizing humanistic practices and embracing their roles as feature identifiers, where health IT systems alone are insufficient.* The physician has historically performed these roles in the absence of health information technology until recent decades, and the art of medicine, a part of the therapeutic patient-physician relationship, remains one of the core oaths of clinical practice. Nonetheless, the ready quantity, availability and reliability of available health data contributing to decision-making has grown significantly. As health information technologies evolve, the predominant focus of the physician must still be to embody the role of healer, yet also evolve as a health data scientist for each patient, collecting data and identifying vital features of a patient's life experience in clinical context. Physicians may elicit and incorporate patient preferences into shared care decisions in ways that current health IT cannot. Human health IT innovations should be designed accordingly, thereby enhancing physicians' ability to focus attention on salient features for the benefit of the patient, especially where vital data cannot be elicited or provided by IT alone.
3. *Human health IT must automate routine or tedious tasks in order to enable high-quality care.* Human health IT should act as a supportive patient care team member, whose role is to facilitate the roles of other team members, and not replace any single entity or relationship on the team. Information systems should automate only the routine or tedious patient care tasks, thereby enabling increased time for dialogue and

essential communication during patient-physician interactions. As human health IT evolves and becomes more intelligent, may acquire additional responsibilities of increasing complexity, yet still maintain a predominantly supportive role that facilitates efficient and high-quality patient-physician interactions.

4. *Human health IT must promote only humanistic patient care.* Enhanced interpersonal continuity of care improves care outcomes [29] and patient satisfaction [30], and is desired by patients, particularly patients with complex chronic conditions [31]. Alternative technologies newly applied to healthcare offer previously infeasible opportunities for patients to access care, for example, via telehealth platforms, mobile apps, and other modalities. Promising applications for specific patient populations and settings [32–34], but further evaluations are needed to identify optimal clinical circumstances where such technology applications maximize benefits and minimize risks [35;36]. Additionally, there may be significant ethical implications for the patient-physician relationship when technology is misapplied, resulting in a loss of the perceptual richness of face-to-face interactions [37]. This may be a limitation to such applications, particularly if current health IT prioritizes access to care at the expense of continuous patient-physician relationships, similar to other health care delivery models that may fragment and increase costs of care [38]. Human health IT must maintain focus on humanistic, continuous care.

Conclusion

Health information technology has evolved rapidly and will inevitably transform the science of clinical medicine. However, the imperative to maintain humanism, or the *art* of medicine, in practice has never been greater, and we posit that human health IT, not simply health IT, is necessary during this new age of digital medicine. Human health IT embodies not only health IT, but the core principle that IT must be designed to support and enhance humanism in medicine. Applied human health IT can automate routine patient care tasks, enabling physicians to apply their vital and finite cognitive resources toward high-value and as yet non-computable tasks, such as addressing human dimensions of care, building rapport with patients, maintaining continuity, and developing a rich and valuable understanding of the medical and psychosocial aspects of each patient's health status. Human health information technology, the next iteration of health IT, must enable humanistic clinical practice and adhere to this high standard, or the digital health revolution dissolve into a biopsychosocial island of future medicine.

Acknowledgements

Views expressed are those of the authors and not necessarily those of the Department of Veterans Affairs or any other affiliated organization. E.L. established the forum, organized meetings and agendas, facilitated discussion, and invited participants and speakers. T.I. assisted with organizing meetings, scribed notes from discussions, and developed initial and final versions of the manuscript. All authors participated in discussions and contributed revisions for the manuscript.

Authors thank Amir Rubin, the former CEO of Stanford Health Care, for his sponsorship of E.L. two-year sabbatical, and providing resources for forum discussions in Year 1. The

authors also thank Year 1 forum participants, including invited guest speakers. Special thanks also to M.G. for hosting discussions in Year 2.

References

- [1] Centers for Disease Control and Prevention, "National Ambulatory Medical Care Survey," *National Ambulatory Medical Care Survey*, 2010. [Online]. Available: <http://www.cdc.gov/nchs/ahcd.htm>. [Accessed: 17-Sep-2015].
- [2] K. S. Yarnall, K. I. Pollak, T. Ostbye, K. M. Krause, and J. L. Michener, "Primary care: is there enough time for prevention?," *Am J Public Heal.*, vol. 93, no. 4, pp. 635–641, 2003.
- [3] T. Ostbye, K. S. H. Yarnall, K. M. Krause, K. I. Pollak, M. Gradison, and J. L. Michener, "Is there time for management of patients with chronic diseases in primary care?," *Ann. Fam. Med.*, vol. 3, no. 3, pp. 209–14, 2005.
- [4] P. E. Kummervold and J. A. K. Johnsen, "Physician response time when communicating with patients over the internet," *J. Med. Internet Res.*, vol. 13, no. 4, pp. 4–8, 2011.
- [5] S. M. Erickson, B. Rockwern, M. Koltov, and R. McLean, "Putting Patients First by Reducing Administrative Tasks in Health Care: A Position Paper of the American College of Physicians," *Ann. Intern. Med.*, Mar. 2017.
- [6] L. S. Sulmasy, A. M. López, and C. A. Horwitch, "Ethical Implications of the Electronic Health Record: In the Service of the Patient," *J. Gen. Intern. Med.*, 2017.
- [7] T. H. Payne et al., "Report of the AMIA EHR 2020 Task Force on the Status and Future Direction of EHRs," *J. Am. Med. Informatics Assoc.*, pp. 1–11, 2015.
- [8] B. Middleton et al., "Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA," *J. Am. Med. Informatics Assoc.*, vol. 20, no. E1, pp. E2–E8, 2013.
- [9] M. Tai-Seale et al., "Electronic Health Record Logs Indicate That Physicians Split Time Evenly Between Seeing Patients And Desktop Medicine," *Health Aff.*, vol. 36, no. 4, pp. 655–662, Apr. 2017.
- [10] C. Sinsky et al., "Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties," *Ann. Intern. Med.*, vol. 165, no. 11, pp. 753–760, Dec. 2016.
- [11] J. E. Siegler, N. N. Patel, and C. J. Dine, "Prioritizing Paperwork Over Patient Care: Why Can't We Do Both?," *J. Grad. Med. Educ.*, vol. 7, no. 1, pp. 16–8, Mar. 2015.
- [12] M. W. Friedberg et al., "Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy," Santa Monica, CA, 2013.
- [13] R. El-Kareh et al., "Trends in primary care clinician perceptions of a new electronic health record," *J. Gen. Intern. Med.*, vol. 24, no. 4, pp. 464–8, Apr. 2009.
- [14] V. Khosla, "Technology will replace 80-percent of what doctors do," *Fortune*. [Online]. Available: <http://fortune.com/2012/12/04/technology-will-replace-80-of-what-doctors-do/>. [Accessed: 17-Sep-2015].
- [15] W. T. Branch et al., "The patient-physician relationship: Teaching the human dimensions of care in clinical settings," *JAMA*, vol. 286, no. 9, pp. 1067–1074, 2001.
- [16] P. F. Weissmann, W. T. Branch, C. F. Gracey, P. Haidet, and R. M. Frankel, "Role modeling humanistic behavior: learning bedside manner from the experts," *Acad. Med.*, vol. 81, no. 7, pp. 661–667, 2006.
- [17] J. S. Ash, M. Berg, and E. Coiera, "Some unintended consequences of information technology in health care: the nature of patient care information system-related errors," *J. Am. Med. Inform. Assoc.*, vol. 11, no. 2, pp. 104–112, 2004.
- [18] G. L. Engel, "The need for a new medical model: a challenge for biomedicine," *Science*, vol. 196, no. 4286, pp. 129–136, 1977.
- [19] J. L. Jameson and D. L. Longo, "Precision medicine—personalized, problematic, and promising," *N. Engl. J. Med.*, vol. 372, no. 23, pp. 2229–34, 2015.
- [20] IOM (Institute of Medicine), *Capturing Social and Behavioral Domains in Electronic Health Records: Phase I*. The National Academies Press, 2014.
- [21] V. J. Felitti et al., "Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study," *Am. J. Prev. Med.*, vol. 14, no. 4, pp. 245–58, May 1998.
- [22] S. R. Greysen and A. S. Detsky, "Understanding the Value of Continuity in the 21st Century," vol. 3, no. 2, pp. 4–6, 2015.
- [23] J. Chi and A. Verghese, "Improving communication with patients: learning by doing," *JAMA*, vol. 310, no. 21, pp. 2257–8, Dec. 2013.
- [24] C.-J. Hsiao and E. Hing, "Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. NCHS data brief, no 143," Hyattsville, MD, 2014.
- [25] D. F. Sittig and H. Singh, "A new sociotechnical model for studying health information technology in complex adaptive healthcare systems,"

- Qual Saf Heal. Care*, vol. 19, pp. 168–174, 2010.
- [26] M. J. Ball, D. E. Garets, and T. J. Handler, “Leveraging Information Technology Towards Enhancing Patient Care and a Culture of Safety in the U.S,” *Methods Inf. Med.*, vol. 42, no. 5, pp. 503–8, 2003.
 - [27] D. M. Zulman, S. M. Asch, S. B. Martins, E. a Kerr, B. B. Hoffman, and M. K. Goldstein, “Quality of Care for Patients with Multiple Chronic Conditions: The Role of Comorbidity Interrelatedness,” *J. Gen. Intern. Med.*, Oct. 2013.
 - [28] C. M. Boyd, J. Darer, C. Boulton, L. P. Fried, L. Boulton, and A. W. Wu, “Clinical Practice Guidelines and Quality of Care for Older Patients,” *JAMA*, vol. 294, no. 6, pp. 716–724, 2005.
 - [29] J. W. Saultz and J. Lochner, “Interpersonal continuity of care and care outcomes: a critical review,” *Ann. Fam. Med.*, vol. 3, no. 2, pp. 159–66, 2005.
 - [30] J. W. Saultz and W. Albedaiwi, “Interpersonal continuity of care and patient satisfaction: a critical review,” *Ann. Fam. Med.*, vol. 2, no. 5, pp. 445–51, 2004.
 - [31] G. Rubin, “Do English patients want continuity of care, and do they receive it?,” *Br. J. Gen. Pract.*, vol. 62, no. 602, p. 462, 2012.
 - [32] K. Day, S. Millner, and H. Johnson, “How Nurses Use Telehealth to Support Health Transitions of Older Adults,” *Stud. Health Technol. Inform.*, vol. 231, pp. 23–30, 2016.
 - [33] V. Wade, J. Karon, A. Elshaug, and J. Hiller, “A systematic review of economic analyses of telehealth services using real time video communication,” *BMC Health Serv. Res.*, vol. 10, no. 1, pp. 1–13, 2010.
 - [34] J. Polisena, D. Coyle, K. Coyle, and S. McGill, “Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations,” *Int. J. Technol. Assess. Health Care*, vol. 25, pp. 339–349, 2009.
 - [35] M.-P. Gagnon and R. E. Scott, “Striving for evidence in e-health evaluation: lessons from health technology assessment,” *J. Telemed. Telecare*, vol. 11 Suppl 2, no. 28, pp. S34–S36, 2005.
 - [36] R. E. Scott, K. Coates, and G. Frank, “The value of an evaluation framework for telehealth initiatives,” *Stud. Health Technol. Inform.*, vol. 64, pp. 39–45, 1999.
 - [37] K. Bauer, “Cybermedicine and the moral integrity of the physician-patient relationship,” *Ethics Inf. Technol.*, vol. 6, no. 2, pp. 83–91, 2004.
 - [38] J. S. Ashwood, M. Gaynor, C. M. Setodji, R. O. Reid, E. Weber, and A. Mehrotra, “Retail Clinic Visits For Low-Acuity Conditions Increase Utilization And Spending,” *Health Aff. (Millwood)*, vol. 35, no. 3, pp. 449–55, Mar. 2016.

Address for correspondence

Correspondence to Tiffany I. Leung, MD, MPH, Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands.
E-mail: t.leung@maastrichtuniversity.nl

Global eHealth, Social Business and Citizen Engagement: A Natural Convergence?

Siaw-Teng Liaw^{a, b, c, d}, Alvin Marcelo^{d, e}, Padmanesan Narasimhan^a,
Md Mahfuz Ashraf^a, Pradeep Ray^{a, d}

^a UNSW Medicine School of Public Health & Community Medicine

(including WHO Collaborating Centre on eHealth & Yunus Social Business Health Hub), Randwick, NSW, Australia

^b Ingham Institute of Applied Medical Research, Liverpool, NSW, Australia

^c SW Sydney Local Health District, Liverpool, NSW, Australia

^d Asia eHealth Information Network

^e University of the Philippines, Manila

Abstract

This paper draws on the vision, mission and experience with the WHO Collaborating Centre on eHealth (WHOCC-eHealth) and Yunus Social Business Health Hub (YSBHH) based at UNSW Australia, and the Asia electronic Health Information Network (AeHIN). Global eHealth aims to provide equitable access to ICT and health care, particularly to the poor, vulnerable and disadvantaged. Social business aims to solve social and economic problem. Its best known product is microcredit financial services for the poor which are small loans that enable them to “produce something, sell something, earn something to develop self-reliance and a life of dignity”. Citizen engagement and community participation is integral to both constructs within the context of global partnerships for Integrated People-Centred Health Services (IPCHS) and Sustainable Development Goals (SDGs). The eHealth dimension is consumer health informatics, social media, mHealth and the Internet of Things. The convergence is multidimensional, mutually beneficial and requires good governance and leadership.

Keywords:

Commerce; Community Participation; Goals

Introduction

The World Health Organization Collaborating Centre (WHOCC) in eHealth was established in The University of New South Wales (UNSW) Medicine in 2013, with its designated activities being evidence-based evaluation, assessment of eHealth and scoping eHealth solutions[1], including the Internet of Things (IoT) [2]. The Asia eHealth Information Network (AeHIN), a group of eHealth advocates in the Asia-Pacific region with an intent on using the peer learning approach to solve their eHealth challenges, is a longstanding collaborator (<http://aehin.org/Home.aspx>). The UNSW Yunus Social Business Health Hub (YSBHH) was established in 2015 to establish, conceive, and promote social business eHealth initiatives.

The scope includes implementation and evaluation of integrated information systems and data, mobile health (mHealth) and working towards an IoT for health. The WHO and the International Telecommunication Union (ITU) sponsored National eHealth Strategy Toolkit is a guiding document that promotes seven strategies for successful implementation of eHealth programs [3]. The vision is global partnerships for Integrated People-Centred Health Services

(IPCHS) [4], United Nations Millennium Development Goals (MDGs) [5, 6], Sustainable Development Goals (SDGs) [7] and health and eHealth workforce [8].

With more than 400 million people globally lacking access to essential health care, the SDGs remain aspirational, like the MDGs. To achieve universal health coverage and equitable access to timely health services, the IPCHS Framework proposes five critical shifts that need to happen (Figure 1): Coordinating services within and across sectors; Re-orienting the model of care; Strengthening governance and accountability; Empowering and engaging people; and Creating an enabling environment.

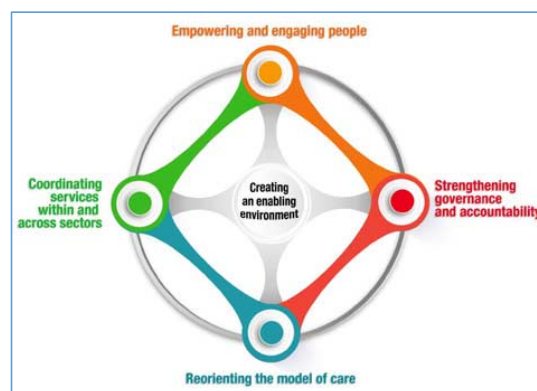


Figure 1 – Five critical shifts required to achieve timely Integrated People Centred Health Services (©WHO 2016)

Developing more integrated people-centred care systems has the potential to generate significant benefits for the health and health care of all people. There is no perfect combination nor a “one size fits all” solution. The right solution will depend on a country’s unique context and needs, as well as local considerations [4].

The MDGs aim to eradicate extreme poverty and hunger, achieve universal primary education, promote gender equality and empower women, reduce child mortality, improve maternal health, combat HIV/AIDS, malaria, and other diseases, ensure environmental sustainability and develop a global partnership for development. The 17 SDGs replaced the MDGs in 2016, with goals relevant to health and well-being:

- SDG#3: Ensure healthy lives and promote well-being for all, at all ages;

- SDG#8: Decent, full and productive work and economic growth;
- SDG#9: Sustainable industry, innovation and infrastructure;
- SDG#10: Income equality within and among countries;
- SDG#11: Sustainable cities and communities, inclusive, safe, resilient and sustainable;
- SDG#12: Responsible and sustainable consumption and production patterns

Social business aims to alleviate social and economic problems caused by poverty, poor health, unhealthy food, smoking, alcohol, gambling, risky behavior, unemployment, poor literacy and other social determinants of health. Social business is not a charity [9]. It is a non-dividend, non-profit business entity [10], but like for-profit businesses, it has to be sustainable. Costs must be recovered, but with the profit-maximization principle replaced by the social-benefit principle. Social business and SDGs complement each other, the SDG identifies a socioeconomic area for development while the social business model of cost-effective execution works to improve that area sustainably [11].

An application of social business methodology to financial services is the Grameen Bank or “*village bank for the poor*”, which makes “microcredit” or tiny loans to poor people. It is based on trust, with no collateral required or legal documents involved! Grameen Bank currently has nine million borrowers, who are “owners” of the bank and lends out over one and a half billion US dollars each year. Almost all (97%) of borrowers are women, who usually use the loan start a business to earn a livelihood. As Prof Yunus describes it, “*by producing something, selling something, earning something, she starts to develop self-reliance and a life of dignity*” [12].

Successful social businesses in health included selling vegetable seeds at affordable prices to make vegetable growing easy for the citizenry. This business has become the largest seed retailer in Bangladesh and more importantly, is associated with a marked reduction of night blindness, a common disease among the poor children in rural Bangladesh. Malnutrition is being addressed by a joint venture in 2005 with a for-profit global company, Danone, to establish a social business to manufacture an affordable fortified yogurt for poor families. This successful social business continues as Danone’s growing corporate social responsibility activities.

A link to eHealth began with the Grameen Village Phone Program. Started in 1997, this provided a good income-earning opportunity to more than 210,000, mostly women, Village Phone operators living in rural Bangladesh through facilitating universal access to telecommunication services by the poor in remote, rural areas. The phone was used mainly for financial discussions and social calls with family and relatives living and working in urban areas, resulting in real savings through avoidance of and reduction in trips to the city [13].

Current social business technologies evolved from adoption of emerging information and communication technologies to affect innovations in social business and the social sector. In 2006, Grameenphone initiated *HealthLine 789* for its ten million subscribers, who are charged thirty-eight cents US dollars per call for five minutes. A range of medical information facilities, (e.g., SMS-based laboratory reports), emergency and ambulance services, and real-time medical consultations are provided via mobile phones. A panel of skilled health professionals is available 24/7 through the physician’s interface and support is provided by a back office and network manager [14, 15, 16].

Increasing smartphone penetration [17], strong users and patient demand for mobile phone apps are strong drivers for mHealth [18]. Health professionals may resist this potential power shift to patients and the community. Regulations of the mHealth industry do not appear to be a barrier but, uncertainty exists around the lack of data security and standards. A general barrier is “*discoverability*”, where it is difficult to discover the required app from among the 100 thousand plus mHealth apps available online.

Community readiness for eHealth is important globally, as well as in rural Bangladesh[19,20], Community members, leaders and healthcare providers would use mHealth tools and services. However, awareness of existing services is low, especially among the poor and less educated. While face-to-face consultations are preferred, the community is attracted by the timely access to qualified healthcare providers, time savings and lower costs associated with mHealth. Low literacy, lack of English language proficiency, lack of trust and technological incapability are barriers to mHealth use [21]. However, a sense of ownership, evidence of utility, a positive attitude and intention to use mHealth were drivers of adoption of mHealth services. Implementation strategies must emphasize gaining the trust, training and support of users. This requires citizen engagement to inform and empower consumers and ensure transparency and accountability.

The key construct in citizen engagement is public participation. Unlike public communication to inform the public, public participation is characterized by a two-way flow of communication in an iterative fashion. It involves the public in collaborative ways and emphasises empowerment. However, barriers exist including poverty and a decreased sense of worth, especially among those with disabilities and disadvantages such as extremes of age, female gender and belonging to minority groups [22].

Our key assertion is that global eHealth implementation and evaluation requires social business strategies, targeted at both clinical and population issues, underpinned by citizen engagement if they are to succeed in improving global health through the IPCHS Framework (Figure 1) and National eHealth Strategy Toolkit.

Methods

A literature search, using “social business”, “eHealth”, “electronic health”, and “health” with MEDLINE (1966-2016), EMBASE (1974-2016) and SCOPUS (1960-2016), found only one paper that met all criteria [23] and 44 papers matched on outcome criteria. Only relevant papers were used to guide the critical analysis of the WHOCC-eHealth and YSBHH activities and critical reflection to focus key principles into a conceptual framework for a convergent research and development program on eHealth and social business.

Findings

The dimensions of the sociotechnical framework to converge eHealth [3] and social business [10] concepts and strategies include: 1. integrated infrastructure and building blocks; 2. collaborative eHealth activities with a citizen focus; 3. citizen engagement and collaboration; 4. measurement and evaluation of citizen-centred process and impacts across the relevant SDGs for the individual, family, community at facility, district, regional and national levels.

Transparent governance, management and leadership across the four dimensions is essential to manage the change needed

to achieve the Triple Aim of better health, better care and cost-effectiveness [24].

1. Convergent building blocks and infrastructure

These include standards and benchmarks for data, information systems, software applications, standard operating procedures, clinical and managerial protocols, change management and governance structures to support learning health organizations and networks to achieve the SDGs for the country. Convergent activities with potential benefit include:

- Establishing the infrastructure and tools to support an user-centered Internet of Things [25].
- Collaborative projects with the Australian Collaborative Research Centre on Spatial Information (CRC-SI) to geocode digital data repositories collected in integrated health neighborhoods [26].
- Building a robust and trusted eHealth infrastructure [27, 28] to support the implementation and monitoring of programs to achieve the SDGs and, previously, the MDGs.
- Building the eHealth workforce capacity through professional exchanges, education and training [29, 30].

2. Convergent citizen-centered eHealth tools & things

These include software applications to contribute to the IoT. Legally, "things" should be seen as an "inextricable mixture of hardware, software, data and service" [25]. Examples of "things" may be wireless devices for detecting and/or monitoring of activities and physiological functions, environment quality, food safety, pathogen activity or functioning of smart homes. This IoT framework applies to eHealth activities such as:

- Cloud-based mHealth systems for disaster management in Indonesia [31];
- mHealth for the primary care of cancer patients in Sydney Australia [32];
- Use of tablets for health checks of independent-living elders, within the Silvercare model where a young, retired person supported up to ten elderly people in their neighborhood with the Indian Aboriginal health agencies [33].

3. Citizen engagement, collaboration and convergence

A digital citizen uses information and communication technology (ICT) to engage in society, politics, and government participation. The key concept here is that it is most efficient and effective at the local community and facility levels. However, the governance and accountability needs to link upwards to meso- and macro-organizations at regional, jurisdictional and national levels. Citizen engagement is an evolving dimension of the WHOCC-eHealth and YSBHH. Activities include:

- Assessing community readiness for mHealth with developing countries [19, 20, 34].
- Assessing market adoption, cost, maturity and user acceptance of robotic mHealth services for vulnerable groups with European Union partners [35].
- Education and training of students from high schools, universities and communities [36].

- Collaborative activities with relevant communities and stakeholders organizations,

4. Convergence and integrated health services & information in local neighborhoods

Digital data from electronic health records (EHRs) of health services in local neighborhoods form the core of any eHealth and health services research and development program. The local neighborhood with its local hospital, general practice and primary care services is the most relevant and logical unit of analysis (denominator) for health services and population health research. Data collected as part of routine clinical practice in EHRs and health information systems (HISs) will form the core longitudinal data source, supplemented by more specific quantitative and qualitative data collection methods at specific time points.

We have defined an Integrated Health Neighborhood (IHN) as a referral network of health services across primary and secondary care, supported by an informatics infrastructure and record linkage across clinical and population health information systems, traditional research data sets, social media and personalized appliances [26]. We have developed and validated tools to assess and manage the quality of data collected and stored in HISs during routine clinical or managerial practices [37-40]. We have also conducted research into natural language processing to improve quality of routinely collected health and social data [41-44].

This evaluation methodology, built around networks of IHNs, is the most logical patient and local community centered approach to collecting data to monitor the implementation, progress and impacts of health care interventions. The IHN approach also enables comparative effectiveness research across IHNs and communities to understand variations in quality of data and care, cost-effectiveness of eHealth in improving self-management, equity and access to health care and social capital in local communities.

We have consulted widely and developed models of data quality management and governance needed to ensure ethical and innovative use of the data collected through this informatics infrastructure [39-41,45-48].

This is an invaluable asset to enable ongoing monitoring of safety and quality of care and achieve the Triple Aim [24] of eHealth and health services, i.e., achieving optimal care, cost and health.

Discussion

Global eHealth research and development, with a social business and citizen engagement approach to achieve integrated people centered health services, is logical and sustainable. It gives meaning to eHealth as mechanisms to achieve access, equity, safety, quality and continuity and comprehensiveness of care of both communicable and non-communicable diseases. Individual and community readiness to adopt and use eHealth tools must be assessed within a comprehensive sociotechnical framework and the context in which they work to achieve healthy behavior and patient outcomes. Finally, open and transparent governance, management and leadership is essential. A range of governance structures exist, but the aim is to extend conventional thinking from simply a government-led program to citizen-led and government-enabled eHealth systems.

Citizen engagement is not easy! However, this is exactly where the social business methodology and approach come into its own. Like the Village Phone project, the community can have a stake in the eHealth program as employees or investors through microcredit for eHealth implementation, training and support projects within the local community. The social business paradigm has shown that it can work.

Integrated people centered health care requires integrated data and information systems (interoperability) to support it and integrated health professionals and services [49, 50] to deliver it to individuals, families and communities. The same data collected during the provision of care should be used to monitor the process and impact on the SDGs [51]. This requires the integration of data from facilities and micro-organizations in health neighborhoods to networks of those managed and governed by meso-organizations at regional levels and macro-organizations further up the hierarchy in the health system.

Most national health systems have this health neighborhood based hierarchical organizational structure. However, the collection, management and stewardship of the data and information is not organized, managed or governed to maximize the benefits of this organizational structure. This need to collect and use meaningful information to monitor progress in the achievement of the SDGs may be a catalyst for developing countries to lead the way for a cost-effective means to organize national HISs to use routinely collected digital data to support health care planning, delivery, evaluation and clinical re-design to iteratively improve the process and impacts of integrated people centered health services.

Finally, mechanisms for countries to share and compare experiences and lessons from the field can trigger and facilitate international learning, which will shorten the transformations in capacity and infrastructure necessary for successful citizen-led relevant social businesses and useful eHealth systems.

Acknowledgements

Nobel Laureate Professor Muhammad Yunus for advice on Social Business; UNSW Sydney Australia, SW Sydney Local Health District and Ingham Institute for Applied Medical Research for infrastructure funding; Australian Primary Health Care Research Institute, HCF Research Foundation, National Prescribing Service and SW Sydney Primary Health Network for project funding; Asia eHealth Information Network and WHO (Western Pacific Region) for regional support; and our collaborators including: Norwegian Centre for Telemedicine; WHO CC (NST); EU partners (www.victoryahome.com) & Technical University of Eindhoven; & partners in European Union (www.au2eu.org); NSW Department of Technology & Industry and partners; Smart Services Collaborative Research Centre (CRC) and industry partners; CRC-Spatial Information and industry partners; University of Sunshine Coast; Griffith University; Aged Care Industry IT Council Australia (AIITC); Dhaka University and ICDDR,B Bangladesh; Yunus Centre & partners in Japan, Thailand, UK, Germany and Australia; IIT-Kharagpur, IIM-Calcutta, India; Nagasaki and Hiroshima Universities, Japan; Fudan and Shanghai Jiao Tong Universities China; Institute of Technology-Bandung, Universitas Gadjah Mada Indonesia; IEEE Communication Society (Healthcom) and IEEE Society for Social Implications of Technology; University of Sussex, UK; University of Colorado, USA.

References

- [1] OECD. ICTs & the Health Sector: Towards Smarter Health and Wellness Models. OECD Publishing; 2013.
- [2] UK Government Office for Science. Internet of Things: making the most of the second digital revolution. London: Government Office for Science; 2014.
- [3] World Health Organisation & International Telecommunication Union. WHO-ITU: National eHealth Strategy Toolkit. Geneva: WHO & ITU; 2012.
- [4] WHO Secretariat. Framework on integrated, people-centred health services. Geneva: WHO; 2016 15 Apr. Report No.: A69/39.
- [5] United Nations. The Millennium Development Goals Report 2015. New York: United Nations; 2015 Jul 01.
- [6] United Nations MGTF. Millenium Development Goal 8: Taking Stock of the Global Partnership for Development. New York: United Nations; 2015.
- [7] United Nations. Sustainable Development Goals: 17 goals to transform our world 2016 [The seventeen Sustainable Development Goals (SDGs) are our shared vision of humanity and a social contract between the world's leaders and the people.]. Available from: <http://www.un.org/sustainabledevelopment/blog/2015/12/sustainable-development-goals-kick-off-with-start-of-new-year/>.
- [8] World Health Organisation. Working for health and growth: investing in the health workforce. Report of the High-Level Commission on Health Employment and Economic Growth. WHO Press, 20 Avenue Appia, 1211 Geneva 27, Switzerland: WHO(2016). ISBN 978 92 4 151130 8.
- [9] Yunus M, Moingeon B, Lehmann-Ortega L. Building Social Business Models: Lessons from the Grameen Experience. *Long Range Planning*. 2010;43(2–3):308–25.
- [10] Yunus M. Creating a world without poverty. Social business and the future of capitalism. New York, USA: *Public Affairs*; (2007).
- [11] Yunus M, Dalsace F, Menascé D, Faivre-Tavignot B. Reaching the Rich World's Poorest Consumers. *Harvard Business Review France*. 2015. <https://hbr.org/2015/03/reaching-the-rich-worlds-poorest-consumers>
- [12] Dowla A. In credit we trust: Building social capital by Grameen Bank in Bangladesh. *The Journal of Socio-Economics*. 2006;35(1):102–22.
- [13] Richardson D, Ramirez R, Haq M. Grameen Telecom's Village Phone Programme in Rural Bangladesh: a Multi-Media Case Study. TeleCommons Development Group, Canadian International Development Agency; 2000 17 Mar.
- [14] Akter S, Ray P, D'Ambra J. Continuance of mHealth services at the bottom of the pyramid: the roles of service quality and trust. *Electron Mark*. 2013;23(1):29–47.
- [15] WHO. mHealth New horizons for health through mobile technologies Geneva, Switzerland: WHO; 2011.
- [16] Al Mamoon I, Khan S. Performance Analysis of a Nationwide Telemedicine Call Center. *J Telecommunications*. 2011;8(2):10–3.
- [17] Pew Research Center. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Pew Research Center 2016.
- [18] Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. *J Med Internet Research*. 2016;18(11).
- [19] Khatun F, Heywood A, Ray P, Bhuiya A, Liaw ST. Determinants of readiness to adopt mHealth in a rural community of Bangladesh. *Int J Med Informatics*. 2015.
- [20] Khatun F, Heywood AE, Ray PK, Bhuiya A, Liaw ST. Community readiness for adopting mHealth in rural Bangladesh: A qualitative exploration. *International Journal of Medical Informatics*. 2016;93:49–56.
- [21] Narasimhan P, Kottapalam A, Macintyre R, Bakshi K, Mathai D, Ray P. Reasons for refusal to participate in a m-Health study for Tuberculosis drug adherence in south India. *Indian Journal of Medical Informatics*. 2013;7(2):111–20.
- [22] Sheedy A, MacKinnon MP, Pitre S, Watling J. Handbook on Citizen Engagement: Beyond Consultation. Canadian Policy Research Networks; 2008 March.
- [23] Mink J, Peterson C. MobileODT: a case study of a novel approach to an mHealth-based model of sustainable impact. *mHealth*. 2016;2(4).
- [24] Berwick D, Nolan T, Whittington J. The Triple Aim: care, health and cost. *Health Affairs*. 2008;27(3):759–69.
- [25] Noto La Diega G, Walden I. Contracting for the 'Internet of Things': Looking into the Nest. In: Law Queen Mary School of Law, University of London 2016.
- [26] Liaw S, de Lusignan S. An 'integrated health neighbourhood' framework to optimise the use of EHR data. *J Innovation in Health Informatics*. 2016;23(3):547–54.
- [27] Akter S, D'Ambra J, Ray P. Trustworthiness in mHealth Information Services: An Assessment of a Hierarchical Model with Mediating and Moderating Effects Using Partial Least Squares (PLS). *J Amer Soc Information Science and Technology (JASIST)*. 2011;62(1):100–16.
- [28] Akter S, Ray P. mHealth - an Ultimate Platform to Serve the Unserved. *IMIA Year Book of Medical Informatics*. 2010:75–81.

- [29] Liaw S, Gray K. Clinical health informatics education for a 21st century world. In: Hovenga E KM, Garde S, Cossio CHL (Eds). *Health Informatics: An Overview. Stud Health Tech & Informatics*. 151: IOSPress; 2010.
- [30] Liaw S, Hovenga E, Aust Council of HI Education. *Health Informatics Education in Australia: Strategic Work Plan for 2009-10 and beyond*. Australian College of Health Informatics.; 2009 October.
- [31] Guempana Y, Rabhi F, Lewis J, Ray P, Zhu L, editors. *Mobile Cloud Computing for Disaster Emergency Operations*. IEEE ISTAS2015; 2015 Nov; Dublin, Ireland.
- [32] Lewis J, Ray P, Liaw S-T. Recent worldwide developments in eHealth and mHealth to more effectively manage cancer and other chronic diseases - a systematic review, . *IMIA Yearbook of Medical Informatics*. 2016;11-26.
- [33] Blake J, Ray P. Facilitating Digital Communication in Seniors. *IEEE International Symposium on Technology and Society (ISTAS2016)*; 2016 Oct 21-22; Trivandrum, India.
- [34] Khatun F, Heywood A, Bhuiya A, Liaw ST, Ray P. Prospects of mHealth to improve the health of the disadvantaged in Bangladesh. In: Adibi S, editor. *mHealth multidisciplinary verticals*: Taylor & Francis; 2015.
- [35] Ariani A, Kapadia V, Talaie-Khoei A, Li J, Ray P. Challenges in Seniors Adopting Assistive Robots: A Systematic Review. *Int Tech Management Review*. 2016 (6); 2: 25-36.
- [36] Nathan S, Kemp L, Bunde-Birouste A, MacKenzie J, Evers C, Shwe T. "We wouldn't of made friends if we didn't come to Football United": The impacts of a football program on young people's peer, prosocial and cross-cultural relationships. *BMC Public Health*. 2013;13:399.
- [37] de Lusignan S, Liaw ST, Krause P, et al. Key concepts to assess the readiness of data for International research: Data quality, lineage and provenance, extraction and processing errors, traceability, and curation. *IMIA Yearbook of Medical Informatics*. 2011;112-121.
- [38] de Lusignan S, Liaw ST, Michalakidis G, Jones S. Defining data sets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology driven approach. *BCS Informatics in Primary Care*. 2011;19(3):127-34(8).
- [39] Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. eGEMs (Generating Evidence & Methods to improve patient outcomes). 2016;4(1):Article 18.
- [40] Liaw S, Rahimi A, Ray P, et al. Towards an ontology for data quality in integrated chronic disease: a realist review of the literature. *Int J Med Informatics*. 2013;82(1):10-24.
- [41] Jonnagaddala J, Liaw S, Ray P. Impact of data quality assessment on development of clinical predictive models. *Stud Health Tech & informatics*. 2015(216):1069.
- [42] Jonnagaddala J, Liaw S, Ray P, Kumar M, Chang N, Dai H. Coronary artery disease risk assessment from unstructured electronic health records using text mining. *J Biomed Informatics*. 2015.
- [43] Jonnagaddala J, Liaw S, Ray P, Kumar M, Dai H. HTNSystem: Hypertension information extraction system for unstructured clinical notes. *Technologies & Applications of Artificial Intelligence*. 2014:219-27.
- [44] Jonnagaddala J, Liaw S, Ray P, Kumar M, Dai H, Hsu C. Identification and Progression of Heart Disease Risk Factors in Diabetic Patients from Longitudinal Electronic Health Records. *BioMed Research International*. 2015(10).
- [45] de Lusignan S, Liyanage H, Di Iorio C, Chan T, Liaw ST. Using routinely collected health data for surveillance, quality improvement and research: Framework and key questions to assess ethics, privacy and data access. *J Innov Health Inform*. 2015;22(4):426-32.
- [46] de Lusignan S, Liaw ST, Krause P, Curcin V, Vicente M, Michalakidis G, Agreus L, Leysen P, Shaw N, Mendis K. . Key concepts to assess the readiness of data for International research: Data quality, lineage and provenance, extraction and processing errors, traceability, and curation. *IMIA Yearbook of Medical Informatics*. 2011;112-121.
- [47] Rahimi A, Liaw ST, Ray P, Taggart J, Yu H. Ontological specification of quality of chronic disease data in EHRs to support decision analytics: a realist review. *Decision Analytics*. 2014;1(1):5.
- [48] Taggart J, Liaw ST, Yu H. Structured data quality reports to improve EHR data quality. *International Journal of Medical Informatics*. 2015;84:1094-8.
- [49] Liaw ST, Taggart J, Yu H. EHR-based disease registries to support integrated care in a health neighbourhood: an ontology-based methodology. *MIE2014*; Istanbul, Turkey.
- [50] Liaw ST, Taggart J, Yu H, de Lusignan S, Kuziemsky C, Hayen A. Integrating electronic health record information to support integrated care: practical application of ontologies to improve the accuracy of diabetes disease registers. *J Biomed Informatics*. 2014;52:364-72.
- [51] Safran C, Bloomrosen M, Hammond WE, et al. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. *Journal of the American Medical Informatics Association*. 2007;14(1):1-9.

Address for correspondence

Professor Siaw-Teng Liaw,
Academic General Practice Unit, UNSW Medicine
1 Campbell Street, Liverpool, NSW 2170 Australia
Email: siaw@unsw.edu.au

P-Care BPJS Acceptance Model in Primary Health Centers

Hosizah Markam

Health Information Management Department, Faculty of Health Sciences, Esa Unggul University Jakarta, Indonesia

Abstract

Electronic Medical Records (EMR) are increasingly adopted in healthcare facilities. Recently, implementation failure of electronic information systems is known to be caused by not only the quality of technical aspects, but also the user's behavior. It is known as applying the Technology Acceptance Model (TAM). This research aimed to analyze the acceptance model of *p-care BPJS* in the primary health centers. A total sample of 30 *p-care BPJS* users was drawn by multistage random sampling in which of these 30 primary health centers participated. Data analysis used both descriptive and inferential statistics. In the phase of structural model, it indicated that *p-care BPJS* acceptance model in the primary health centers was formed by Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) through Attitude towards use of *p-care BPJS* and Behavioral Intention to use *p-care BPJS*.

Keywords:

Electronic Health Records; Hospital Information Systems; Models, Theoretical

Introduction

Information and communication technologies (ICTs) have great potential to improve health in both developed and developing countries by enhancing access to health information and making health services more efficient; they can also contribute to improving the quality of services and reducing their cost. Today, most electronic information systems used in the primary health centers only store data or patient aggregate information at the management level. Based on findings from the global eHealth survey by the WHO and the World Bank, the State of Africa and Southeast Asia were the highest (over 90%) individual patient data usage in paper-based format, whereas electronic medical records (EMR) were mostly adopted in developed and developing countries, such as Brazil, China and India [1-4].

An electronic medical record (EMR) is a real-time patient health record with access to evidence-based decision support tools that can be used to aid clinicians in decision-making. The EMR can automate and streamline a clinician's workflow, ensuring that all clinical information is communicated. It can also prevent delays in response that result in gaps in care. The EMR can also support the collection of data for uses other than clinical care, such as billing, quality management, outcome reporting, and public health disease surveillance and reporting. Furthermore, an EMR may contain clinical applications that can act on the data contained within its repository, for example, a clinical decision support system (CDSS), a computerized provider order entry system (CPOE), a controlled medical vocabulary, or a results-reporting system. In general, EMRs are clinician-focused in that they enhance or augment the workflow of clinicians or administrators [1, 3, 5,

6]. Currently, the only EMR concept used in primary health centers in Indonesia is *p-care BPJS*.

P-Care BPJS is Primary-Care BPJS. BPJS is Badan Penyelenggara Kesehatan or National Social Security. In Indonesia, *P-Care BPJS* is commonly known as *p-care BPJS* in Healthcare. *P-Care BPJS* is a web-based patient care application provided by BPJS. It is used for accessing data stored in the BPJS server including patient registration, laboratory, diagnosis, treatment and therapy [7]. Recently, implementation failure of electronic information systems is caused not only by the quality of technical aspects, but also the user's behavior. Applying the TAM may help reduce implementation failure [8, 9]. TAM is an information system theory that models how users come to accept and use technology: the main dependent constructs are behavior intention to use and system usage. Until now in Indonesia, studies examining the individual acceptance of the use of *p-care BPJS* have not been done yet. This study aimed to analyze the technology acceptance model of *p-care BPJS* in the primary health centers.

Methods

This cross-sectional observational study was conducted in February-May 2016.

The study population consisted of users of *p-care BPJS* at primary health centers in the province of East Java. Users of *p-care BPJS* were responsible to operate the *p-care* application in primary health center. A sample size of 30 users of *p-care BPJS* from 30 primary health centers in five districts, namely: Bangkalan; Bondowoso; Lamongan; Malang; Kediri with multistage sampling method, each district was represented by 6 primary health centers and one *p-care BPJS* user from each were included.

According to TAM, this study posited the following seven hypotheses:

H₁: Perceived Ease of Use (PEOU) significantly affects to Perceived Usefulness (PU)

H₂: Perceived Ease of Use (PEOU) significantly affects to Attitude towards Use *p-care BPJS*

H₃: Perceived Usefulness (PU) significantly affects to Attitude towards Use *p-care BPJS*

H₄: Perceived Usefulness (PU) significantly affects to Behavioral Intention to Use *p-care BPJS*

H₅: Perceived Usefulness (PU) significantly affects to Actual *p-care BPJS* Use

H₆: Attitude towards Use *p-care BPJS* significantly affects Behavioral Intention to Use *p-care BPJS*

H₇: Behavioral Intention to Use *p-care BPJS* significantly affects Actual *p-care BPJS* Use

The questionnaire has been prepared in accordance to questions in the TAM [8, 9]. The response scale for all TAM items was a six-point scale, ranging from 1 (Extremely Unlikely) to 6 (Extremely likely). The questionnaire was tested for the validity. The result showed that the validity and reliability was good (the Cronbach's Alpha value was 0.896).

A total of 30 questionnaires were distributed with a 100% response rate. Thirty completed questionnaires were obtained and considered valid. Data analysis used both descriptive and inferential statistics. Inferential statistics was performed with *Structural Equation Modeling* (SEM) by using a SmartPLS 3.0 program consisting of two phases, the measurement model and the structural model [10].

Results

The return rate for the questionnaire was 100%. The characteristics of respondent description in table 1.

Table 1- Characteristics of Respondent

Characteristic of Respondent		Frequency	Percentage (%)
Sex	Male	12	40
	Female	18	60
	Total	30	100
Age (years)	< 20	1	3.3
	20-30	13	43.3
	31-40	16	53.4
	Total	30	100
Education Level	SMA/SMK	6	20
	Diploma	17	57
	Baccalaureate	7	23
	Total	30	100
Type of Education	Health	24	80
	Non-health	6	20
	Total	30	100

Of the 30 respondents, more than one-half were female (60%), and older than 30 years of age (53.4%). Forty three percent were between the ages of 20-30 years of age and younger than 20 years of age (3.3%). Fifty-seven percent of respondents had obtained a 3-year diploma, with the remaining respondents having completed a four-year bachelor's degree and senior high school. Eighty percent passed from health education and the remainder having non-health education.

Construct validity test results with the Smart PLS in the measurement model can be seen in the value of convergent validity and reliability tests with value of composite reliability (CR) and Cronbach's Alpha. In Table 2 the value of the loading factor (convergent validity) was between 0.913-0.949; it can be explained that almost all indicators of loading factor >0.7 and Cronbach's Alpha > 0.6 (0.855-0.933). Thus all the constructs, Perceived Ease of Use, Perceived Usefulness, Attitude toward Use *p-care BPJS*, Behavioral Intention Use *p-care BPJS*, Actual *p-care BPJS* Use have met the acceptable cut-off values of reliability [10, 11].

Table 2- Factor Loadings, Composite Reliability and Cronbach's Alpha

Constructs	Items	Loadings	Composite Reliability	Cronbach's Alpha
Perceived Ease of Use (PEOU)	PEOU 1	0.850	0.949	0.933
	PEOU 2	0.940		
	PEOU 3	0.962		
	PEOU 4	0.781		
	PEOU 5	0.691		
	PEOU 6	0.970		
Perceived Usefulness (PU)	PU1	0.903	0.931	0.910
	PU2	0.902		
	PU3	0.640		
	PU4	0.900		
	PU5	0.831		
	PU6	0.794		
Attitude toward Use <i>p-care BPJS</i> (ATU)	ATU1	0.879	0.944	0.910
	ATU2	0.939		
	ATU3	0.943		
Behavioral Intention Use <i>p-care BPJS</i> (BIU)	BIU	0.948	0.913	0.855
	BIU2	0.911		
	BIU3	0.782		
Actual <i>p-care BPJS</i> Use (AU)	AU1	0.767	0.923	0.876
	AU2	0.962		
	AU3	0.945		

The test results of the inner model consisted of a coefficient parameter path (path coefficient parameter) and the value of R Squared (R^2) in Table 3 and Table 4.

Table 3- Path Coefficient of Variables

Direct and indirect influence between endogenous and exogenous variable	B	Sample Mean (M)	Standard Error	P-values
PEOU → Attitude toward Use <i>p-care BPJS</i> (ATU)	0.521	0.165	0.327	0.041*
PEOU → Perceived Usefulness (PU)	0.892	0.894	0.036	0.000**
PU → Attitude toward Use <i>p-care BPJS</i> (ATU)	0.726	0.591	0.332	0.000**
PU → Actual <i>p-care BPJS</i> Use (AU)	0.108	0.130	0.163	0.507
PU → Behavioral Intention Use <i>p-care BPJS</i> (BIU)	0.260	0.281	0.172	0.130
Attitude toward Use <i>p-care BPJS</i> (ATU) → Behavioral Intention Use <i>p-care BPJS</i> (BIU)	0.657	0.627	0.172	0.000**
Behavioral Intention Use <i>p-care BPJS</i> (BIU) → Actual <i>p-care BPJS</i> Use (AU)	0.808	0.793	0.147	0.000**

Note: *sig=0.05, **sig=0.01)

The result of path analysis in Table 3 shows that actual p-care BPJS Use were significantly influenced by Perceived Ease of Use through Perceived Usefulness, Attitude toward Use p-care BPJS, Behavioral Intention Use p-care BPJS. Actual p-care BPJS Use was not significantly influenced directly by Perceived Usefulness and also on Behavioral Intention Use p-care BPJS was not significantly influenced directly by Perceived Usefulness. A coefficient parameter path of Perceived Ease of Use p-care BPJS to Attitude toward Use p-care BPJS is smaller than through Perceived Usefulness p-care BPJS.

Table 4- R Squared (R^2)

Constructs	R-square (R^2)
Perceived Usefulness (PU)	0.793
Attitude towards Use p-care BPJS (ATU)	0.528
Behavioural Intention to Use p-care BPJS (BIU)	0.716
Actual p-care BPJS Use (AU)	0.790

In table 4, R-squared Actual p-care BPJS Use amounted 0.790, Behavioral Intention to Use p-care BPJS is 0.716 and Perceived Usefulness is 0.793.

- Perceived Usefulness

Perceived usefulness as the extent to which a person believes that using a technology will enhance her or his performance. Perceived Usefulness consists of six items: work more quickly, job performance, increase productivity, effectiveness, makes job easier, and useful. Participants responded likely and extremely likely on all of item work more quickly, job performance, increase productivity, effectiveness, makes job easier, and useful more than 75%.

Path coefficient of perceived usefulness affected Attitude towards Use p-care BPJS by 0.726 and significant p-value of 0.000 (<0.05). This can be explained that Perceived Usefulness user p-care BPJS positive effect and significantly influenced on Attitude toward Use p-care BPJS. Perceived Usefulness users of p-care BPJS had a positive effect on both Behavioral Intention and Actual p-care BPJS Use but were not significantly significant.

- Perceived Ease of Use

Perceived ease of use is the extent to which a person believes that using a technology will be free of effort. Perceived Ease of Use is composed of six items namely easy of learn, controllable, clear and understandable, flexible, easy to become skillful and easy to use. Participants responded likely and extremely likely on all of items easy of learn, controllable, clear and understandable, flexible, easy to become skillful and easy to use more than 79%. Effect of perceived ease of use to the perceived usefulness with path coefficient of 0.892 and significant p-value of 0.000 (<0.05). Perceived Ease of Use positively affected Perceived Usefulness users of p-care BPJS.

- Attitude Toward Use p-care BPJS

Attitude toward Use p-care BPJS is an individual's positive or negative feelings about performing the target behavior. Attitude toward use p-care BPJS is composed of three items have fun, enjoyment and boring. Participants responded likely and extremely likely on all of items have fun, enjoyment and boring more than 82%.

Attitude toward use of p-care BPJS positively affected Behavioral Intention Use p-care BPJS with path coefficient of

0.657 and significant p-value of 0.000 (<0.05). Attitude toward use of p-care BPJS of primary health center had a positive effect on Behavioral Intention Use p-care BPJS.

- Behavioral Intention Use p-care BPJS

Behavioral intention Use p-care BPJS consists of three items intend, predict and plan. Behavioral intention to use p-care BPJS had positive effect on actual p-care BPJS use with the path coefficient of 0.808 and significant with a p-value of 0.000 (<0.05). Behavioral intention to use p-care BPJS positively affected actual p-care BPJS effect on primary health center.

- Actual p-care BPJS Use

Actual p-care BPJS use was significantly influenced by perceived ease of use through perceived usefulness, attitude toward use p-care BPJS, and behavioral intention to use p-care BPJS. Actual p-care BPJS Use was not significantly influenced by perceived usefulness. Behavioral intention to use p-care BPJS was also not significantly influenced by perceived usefulness.

Discussion

This study analyzed the technology acceptance model (TAM) factors influencing p-care BPJS users' actual use in the Primary Health Centers. In particular, the influences of perceived usefulness, perceived ease of use, and attitude towards p-care BPJS on intention to use p-care BPJS are examined. At the same time, hypotheses about the influence among perceived usefulness, perceived ease of use, and attitude towards p-care BPJS use were also tested.

There was a strong effect (0.790) of behavioral intention to use p-care BPJS [11]. This showed that the perceived usefulness, attitude toward use of p-care BPJS and behavioral intention to use p-care BPJS could explain the variance of actual p-care BPJS amounted to 79% and the remaining 21% was influenced by other variables.

Perceived usefulness affected attitude toward use of p-care BPJS, which is in line with previous studies of the acceptance on the Clinical Information Systems among medical staff [12], [13], [14].

Perceived usefulness users of p-care BPJS had a positive effect on behavioral intention to use p-care BPJS but was not significant. Perceived usefulness was not affected by behavioral intention to use p-care BPJS, which is in line with previous studies [15]. Perceived usefulness of p-care BPJS had a positive effect on actual p-care BPJS use, but was also not significant. Perceived usefulness affected actual p-care BPJS use. This is in line with previous studies [15].

Perceived ease of use positively affected on perceived usefulness of p-care BPJS users of the primary health center, which is in line with the TAM as proposed by Davis as well as the previous studies [8][16][17]. Perceived ease of use positively affected attitude towards use of p-care BPJS. Perceived ease of use positively affected attitude toward use of p-care BPJS user of the primary health center, this is in line with the previous studies physician's acceptance of hospital information systems [16], in the Acceptance of Telemedicine Systems [17].

Attitude toward use of p-care BPJS in primary health centers was positively effected by behavioral intention to use p-care BPJS. This research is also consistent with results of previous studies of in the Acceptance of Telemedicine Systems [17], user acceptance of a picture archiving and communication system (PACS) [18].

Behavioral intention to use *p-care BPJS* affected actual *p-care BPJS* use. This is in line with previous studies of user satisfaction and technology acceptance [22] and students acceptance of mobile learning for higher education [23].

Perceived usefulness affected actual *p-care BPJS* use, this is in line with previous studies of end-users' acceptance and use of hybrid library services [25], electronic medical records: TAM, UTAUT and culture [26].

Goodness of Fit (GoF) index of *p-care BPJS* acceptance model in the primary health centers was 0.741. It is good or fits the model. There are three categories in GoF index: small=0.1; medium=0.25; and large=0.36. GoF index is crucial for assessing the global validity of a complex model [27].

The R^2 for the model was 0.790. This means that the diversity of behavioral data of *p-care BPJS* use in primary health centers that can be explained by the construct is 79%, the remaining 21% is explained by other constructs that were not included in the model. The determinant of *p-care BPJS* acceptance model in the primary health centers is perceived ease of use and perceived usefulness.

Conclusions

Five of seven hypotheses tested here were accepted. There was a difference with the original TAM, but in line with previous studies. Actual *p-care BPJS* use was significantly influenced by perceived ease of use through perceived usefulness, attitude toward use of *p-care BPJS*, behavioral intention to use *p-care BPJS*. Actual *p-care BPJS* use was not significantly influenced by perceived usefulness or behavioral intention to use *p-care BPJS*, not was it significantly influenced by perceived usefulness. Effect of perceived ease of use has become a determinant for successful implementation of *p-care BPJS* in the primary health centers. Thus, users of *p-care BPJS* need to be trained on an ongoing basis to keep up to date in understanding the latest version of the program. As this is a preliminary attempt to understand how the system is being accepted, a more comprehensive rigorous research approach can be adopted in the future that incorporates a larger model, such as Unified Theory of Acceptance and Use of Technology (UTAUT) and others that are able to capture external influences. It is with a hope that *p-care BPJS* will evolve and improve to serve the Primary Health Centers as a whole as well as the community through better health care approach using computer based solutions.

Acknowledgements

I would like to thank Esa Unggul University for supporting and funding this study. I also wish to thank to the respondents for the support and help in participating in the study.

References

- [1] World Health Organization (WHO). *Management of Patient Information: Trends and Challenges in Member States: based on the findings of the second global survey on eHealth*. Global Observatory for eHealth Series-Volume 6 (2012)
- [2] Lippeveld, T., Sauerborn, R., Bodart, C. *Design and Implementation of Health Information Systems*. WHO, (2000), Geneva
- [3] Decree of Ministry of Health (MoH) Republic of Indonesia No.192/Menkes/SK/VI/2012 ROADMAP Action Plan of Strengthening on Health Information Systems, available: <http://depkes.go.id/downloads/RoadMapSIK.PDF> accessed December 11th 2015
- [4] United States Department for Health and Human Services. *Electronic medical records*. The Office of the National Coordinator for Health Information Technology, United States Department for Health and

- Human Services. (2012) (http://healthit.hhs.gov/portal/server.pt/community/electronic_medical_records/1219/home/15591) accessed January 7th 2016
- [5] Cilliers L., Flowerday, S.V., Health information systems to improve health care: A telemedicine case study, *SA Journal of Information Management* 15(1), Art. #541, (2013), 5 pages. <http://dx.doi.org/10.4102/sajim.v15i1.541>
- [6] Johnson DL, Spielman CM. *Data and Information Management*. In: LaTour KM, Maki SE, Ed. Third. Health Information Management: Concepts, Principles and Practice. American Health Information Management Association (AHIMA) (2010): Chicago Illinois
- [7] <http://www.bpjs-online.com/mengenal-pcare-bpjs-kesehatan> accessed December 12th 2015
- [8] Davis, F.D., Perceived usefulness, perceived ease of use, and Acceptance of Information System Technology, *MIS Quarterly*, 13 (1989)
- [9] Jogiyoanto HM, Sistem Informasi Keperilakuan. Yogyakarta: Penerbit Andi, 200
- [10] Hair JF, Hult GTM, Ringle CM, Sarstedt M. (2014) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. United Kingdom: SAGE Publication.
- [11] Mustafa, Wijaya. (2012) *Panduan Teknik Statistik SEM & PLS dengan SPSS AMOS*. Yogyakarta: Cahaya Atm Pustaka
- [12] Christors D., Modeling the acceptance of the clinical information systems among medical staff: An extended TAM model. *Journal Biomedical Informatics*, 4 (2011), 553-564
- [13] Dhesi BR, Cheah WL, Perception of End Users on I-Kelahiran in Sabah, Malaysia, *electronic Journal of Health Informatics*; 8 (2014), 19
- [14] Zhu Ge, Sangwan S., Lu T., A new theoretical framework of technology acceptance and empirical investigation on self-efficacy-based value adoption model, *Nankai Business Review International*, 1 (2010)
- [15] Sentosa I, NikMat NK. Examining a theory of planned behavior (TPB) and technology acceptance model (TAM) In internet purchasing Using Structural Equation Modeling. *Journal of Arts, Science & Commerce* 11 (2012), 2231-4172.
- [16] ChenRai-Fu, HsiaoJu-Ling, An investigation on physician's acceptance of hospital information systems: A case study. *International Journal of Medical Informatics* (2012), 810-820
- [17] Djasasbi S, Fruhling AL, Loiacono ET, The influence of affect, attitude and usefulness in the acceptance of telemedicine systems. *Journal of Information Technology Theory and Application (JITTA): A Publication of the Association for Information System* 10 (2009), 41-58
- [18] Bakheet Aldosari, User acceptance of a picture archiving and communication system (PACS) in a Saudi Arabian hospital radiology department. Aldosari, *BMC Medical Informatics and Decision Making* 12 (2012), 44
- [19] Chihung L, I-Chun L, Jinsheng R, Barriers to physicians' adoption of healthcare information technology: An empirical study on multiple hospitals, *Journal Medical System* 36 (2012), 1965-1977
- [20] C.Lin, I-C. Lin, Roan J., Barriers to Physicians' Adoption of Healthcare Information Technology: An Empirical Study on Multiple Hospitals, *Journal Medical System* (36) (2012), 1965-1977
- [21] Jian WS, Shabbir SA, Sanjay PS, Peisan L, Hsu MinH, Ho CH, Li YC, Wen HC. Factors influencing consumer adoption of USB-based Personal Health Records in Taiwan, *BMC Health Services Research*, (2012) 12:277
- [22] Miyamoto M, Kudo S., Lizuka K. Measuring ERP success: Integrated model of user satisfaction and technology acceptance: An empirical study in japan. *International Proceedings of Economics Development and Research* 57 (2013), 86
- [23] Nassuora A.B., Students Acceptance of Mobile Learning for Higher Education In Saudi Arabia, *American Academic & Scholarly Research Journal* 4 (2012).
- [24] SivanarulselvanSachidanandam, Why Physicians Do or Do Not Use Computerized Physician Order Entry Systems: Applying The Technology Acceptance, ProQuest Information and Learning Company, 2007
- [25] Tibenderana P, Patrick O, Ikoja-Odongo J, Wokadala J. Measuring levels of end-users' acceptance and use of hybrid library services, *International Journal of Education and Development using Information and Communication Technology (IJEDICT)*, 6 (2010), 33-54
- [26] Trimmer K, Cellucci, Wiggins C, Woodhouse W. Electronic Medical Records: TAM, UTAUT and Culture, *International Journal of Healthcare Information Systems and Informatics* 4 (2009), 55-68
- [27] Akter S, D'Ambra J, Ray P. An evaluation of PLS based complex models: the roles of power analysis, predictive relevance and GoF index. (2011). Proceedings of the 17th Americas Conference on Information Systems (AMCIS2011) (pp. 1-7). Detroit, USA: Association for Information Systems.

Address for correspondence

Hosizah Markam
Faculty of Health Sciences
University of Esa Unggul

Arjuna Utara Street No.9, Kebon Jeruk, West Jakarta,

Indonesia 11410

Phone: +62-21-5674223, +62-815-13014577

hozisah@esaunggul.ac.id

The Australian Health Informatics Competencies Framework and Its Role in the Certified Health Informatician Australasia (CHIA) Program

Fernando Martin-Sanchez^{a,b,c}, David Rowlands^c, Louise Schaper^c, David Hansen^{c,d}

^a Environmental and Participatory Health Informatics (ENaPHI) Research Group, Division of Health Informatics, Department of Healthcare Policy and Research, Weill Cornell Medicine, New York, NY, USA

^b Health and Biomedical Informatics Centre, University of Melbourne, VIC, Australia

^c Health Informatics Society of Australia

^d eHealth Research Centre, CSIRO, Brisbane, QL, Australia

Abstract

The Certified Health Informatician Australasia (CHIA) program consists of an online exam, which aims to test whether a candidate has the knowledge and skills that are identified in the competencies framework to perform as a health informatics professional. The CHIA Health Informatics Competencies Framework provides the context in which the questions for the exam have been developed. The core competencies for health informatics that are tested in the exam have been developed with reference to similar programs by the American Medical Informatics Association, the International Medical Informatics Association and COACH, Canada's Health Informatics Association, and builds on the previous work done by the Australian Health Informatics Education Council. This paper shows how the development of this competency framework is helping to raise the profile of health informaticians in Australasia, contributing to a wider recognition of the profession, and defining more clearly the body of knowledge underpinning this discipline. This framework can also be used as a set of guidelines for recruiting purposes, definitions of career pathways, or the design of educational and training activities. We discuss here the current status of the program, its results and prospects for the future.

Keywords:

Certification, Professional Competence, Medical informatics.

Introduction

In other countries, certification in health informatics is achieved through university degrees. However, in Australia, only a few universities offer studies related to health informatics or a full undergraduate or postgraduate degree in this area [1]. In addition, health informatics lacks formal recognition in the Australian health workforce [2].

To address this, the Health Informatics Society of Australia (HISA), in collaboration with the Australasian College of Health Informatics (ACHI) and the Health Information Management Association of Australia (HIMAA), launched the *Certified Health Informatician Australasia* (CHIA) certification program in July 2013 at HIC, the major Health Informatics Conference in Australia. The CHIA consists of an online exam, which aims to test whether a candidate has the knowledge and skills that are identified in the competency framework to perform as a health informatics professional.

A competency is the specification of knowledge and skill, and the application of that knowledge and skill, to the standard of performance expected in the workplace. A competency is the smallest unit that can be assessed and recognized [3].

Competencies should be used to set clear expectations for professionals in providing them with direction on what level they need to be performing in the work place. It also provides professionals with an itinerary for development and continuous education, including closing knowledge gaps. Competencies also provide a framework for career growth and improvement in recruitment processes. Finally, competencies can help establish professional opportunities, including horizontal and vertical moves within an organization [4].

The CHIA Health Informatics Competencies Framework provided the context in which the questions for the exam were developed. It was developed by a committee composed of representatives from HISA, HIMAA and ACHI, and included leaders in health informatics academia and the Health IT sector. The core competencies for health informatics, that are tested in the exam, have been developed with reference to similar programs by the American Medical Informatics Association (AMIA), the International Medical Informatics Association (IMIA) and Canada's Health Informatics Association (COACH), and builds on the previous work done by the Australian Health Informatics Education Council.

Methods

The Health Informatics Competencies Framework

Health and Biomedical Informatics (HBI) is a discipline with more than 50 years of existence. Along this period, many sub-specializations have emerged. HBI is considered as the core set of information methods, theories and tools, which are relevant to healthcare, biomedical research, and public health. However, depending on the application domain, different sub-specialties can be identified. The original intention by the CHIA governing bodies was to introduce specialization exams based on segmentation by discipline (e.g. CHIA-NI for nurse informaticians). However, given workforce trends across the globe that show the need for transferability of skills and knowledge across disciplines, we are now considering a tiered structure of increasingly advanced health informatics knowledge that is not discipline specific. The decision and subsequent development of 'CHIA+' will be occurring from late 2017 onwards. To develop a set of highly regarded competencies in health informatics (HI), we needed firstly to refer to the most relevant existing frameworks:

- AMIA – A leading professional and academic organization in biomedical informatics, published the most recent white paper on core competencies in June 2012 [5];
- IMIA – Peak body at the international level in the area of health and biomedical informatics, published revised education Recommendations in 2010 [6]; and

- COACH – Canada has many similarities to Australia in terms of healthcare systems, population and geography. COACH has a good working relationship and strong ties with HISA. They published Health Informatics Professional Core Competencies in 2012 [7].

Each stream from the three organizations were reviewed carefully under the topic areas, compared, mapped and then regrouped. Any repetitions, overlaps, and redundancies were removed, and the new competencies were restructured, ultimately producing the final competency framework.

Motivation to redraft the Health Informatics Competency Framework

- Some competencies are now outdated, especially some from IMIA, which originates from 2007.
- IMIA recommendations are extremely useful for the development of educational curriculum.
- There is a clear distinction between information technology and information science; therefore, we have addressed this in creating two separate competency streams in these areas.
- Management is more recognized as a relevant set of competencies in the framework by COACH.
- In Australia, Health Information Management (HIM) is already defined as a distinct professional jurisdiction by HIMAA's HIM Competency Standards for tertiary and postgraduate sectors, and HIMAA's Coder Units of Competency in the VET sector.
- The consideration of health informatics as a scientific discipline is addressed by the AMIA competencies.
- The human and social context is addressed by AMIA.
- We recognize the work and competency framework completed by the Australian Health Informatics Education Council (AHIEC) [8]. However, as the AHIEC framework is mostly based on the IMIA framework, it was decided that the broader IMIA document would be used.
- It is also recognized that the health information management competencies developed by HIMAA are complementary to the health informatics set of competencies, but it is more fitting to define the relationship between these competencies in consultation with the health information management professional jurisdiction at a future date.
- Although both IMIA and AHIEC frameworks contain different levels of competencies and different specializations streams, we preferred to regulate the varying levels of complexities in the form of knowledge, comprehension, application and analysis type of questions during the exam writing process rather than within the competency itself [9].

After the exhaustive exercise of reviewing the three abovementioned competency frameworks, including the mapping of competencies between each of the organizations, it was deemed appropriate to develop a new competency framework based on the existing frameworks, but with a focus on the Australian healthcare system.

Eligibility criteria

In order to achieve a CHIA qualification, a candidate must first meet the eligibility criteria to register. The eligibility requirements in order to be able to sit the CHIA exam are:

- A degree plus three years of associated experience¹; OR
- Five years of associated experience (no degree required)².

The eligibility of individuals who don't meet the above criteria are reviewed by the CHIA Examination Committee on a case by case basis.

Results

The Competencies Framework

Figure 1 provides a representation of the Health Informatics Competency Framework showing the six competency streams as a proportion of the overall. The CHIA exam covers all 6 areas. The list of competencies is provided in Table 1, but the complete information with descriptions and levels of assessment can be accessed at the following URL: http://www.healthinformaticscertification.com/wp-content/uploads/2016/02/CHIA-competencies-Framework_FINAL.pdf



Figure 1 – Health Informatics Competencies Framework

The CHIA Examination Study Guide

An examination study guide entitled “A Practitioner's Guide to Health Informatics in Australia” has been published by the Health Informatics Society of Australia (HISA) as a comprehensive body of knowledge (+1000 pages) covering competencies tested in the CHIA exam. It is the definitive guide for CHIA candidates.

The Practitioner's Guide has also been developed with other purposes in mind, including orientation for professionals, such as clinicians or ICT professionals new to health informatics and updates for health informaticians, wishing to maintain the currency of their knowledge, irrespective of certification. It focuses explicitly on the needs of practitioners in the field of health informatics in Australia.

- Covers all of the 52 CHIA competencies at the level of competence required (knowledge, comprehension, application or analysis),
- Aimed at helping candidates achieve certification, and
- Provides a valuable professional and workplace resource.

The Guide is published electronically, into 8 sections and is available for purchase³.

Table 1 - CHIA Health Informatics competencies framework

Domain of expertise	Areas of Competence
Information and Communication Technology (Information technology in general, not limited to healthcare, though the principles certainly apply to healthcare)	1.1 Basic knowledge of ICT concepts 1.2 Problem solving through ICT 1.3 Analysis of stakeholder needs along the System Life Cycle 1.4 Selection and use of ICT 1.5 Good practice in System Life Cycle
Health and Biomedical Science (Healthcare systems and practice and basic biomedical science concepts)	2.1 Basic health and biomedical concepts 2.2 Data, information and knowledge in health and biomedicine 2.3 Factors related to health 2.4 Clinical decision making 2.5 Models of care delivery 2.6 Evidence based clinical practice 2.7 Health administration and health services research 2.8 Epidemiology and basic health research skills 2.9 Clinical language and vocabulary 2.10 Professional roles & resources in health organizations
Information Science (Information systems in general, not limited to healthcare, though the principles certainly apply to healthcare)	3.1 Applicable Mathematical concepts 3.2 Basic knowledge of IS concepts 3.3 Information theories 3.4 Quality principles across the IS life cycle 3.5 Realization of benefits from I.S. 3.6 Attributes & limitations of data & information 3.7 Data analysis and visualisation 3.8 Identification of gaps in data sources
Management Science (Governance and management of systems development, change management, business practices and organizational strategy at all levels)	4.1 Project & change management. 4.2 Alignment of IS with organizational strategies 4.3 Information cultures and learning within organizations 4.4 Good practice in process engineering 4.5 Risk management
Core Principles and Methods (Health informatics theoretical foundations, practice and applications)	5.1 History of HBI & analysis of related literature 5.2 Theories of HBI 5.3 Conceptual frameworks in HBI 5.4 Knowledge representation in HBI 5.5 Governance of IS in HC 5.6 IS to support patients 5.7 Electronic Health Record 5.8 Informatics in support of education & research 5.9 Interfacing & patient identification 5.10 Decision support systems 5.11 Architectures of Health IS 5.12 Interoperability & HI standards 5.13 Integration of clinical data & associated risks 5.14 Clinical safety & IS 5.15 Value of IS & adoption 5.16 Informatics for participatory health 5.17 New data sources & emerging technologies 5.18 E-health applications and solutions 5.19 Knowledge translation in health 5.20 Areas of specialisation in HBI
Human and Social Context (Human and social context related to healthcare and the systems of healthcare including issues of clinical practice, consumers and legal requirements)	6.1 Technology & social aspects 6.2 The relevance of ethical & legal issues for health informatics 6.3 Policies, principles & guidelines for HI management 6.4 Usability & human factors

The CHIA Exam

Becoming a CHIA involves sitting and passing the CHIA exam. The exams covers 6 competency areas and 52 competencies, and includes 104 multiple choice questions (2 questions on each competency at various levels of difficulty). The exam is completed online, and the duration is 2.5 hours maximum. Exam questions are selected at random (i.e. students re-sitting the exam will not be given the same questions as the first time they sat the exam). The exam is managed by an Examination Committee comprising of senior Australasian health informaticians.

Periodic Renewal of Certification

Periodic renewal of the certification is required to maintain certified status. Initial certification or renewal of certification is valid for 3 years. Renewal is required to maintain the CHIA certification. The renewal requirements are currently under review and are likely to be: Submit a recertification journal that documents a minimum level of ongoing professional development or take the exam again.

Discussion

Comparative analysis of competencies in health informatics

We provide the comparative analysis of groups and competencies (Table 2) for a breakdown of the different areas of competencies covered.

Results so far

The official program launch took place on July 16, 2013 at the Health Informatics Conference in Adelaide, in conjunction with a workshop and education session. The first exams took place in December 2013, where 415 candidates have registered to sit the exam, 57 have forfeited (not attempted at all), 32 are pending a first attempt, 84 have failed, and 242 have passed.

Candidates come from a range of organizational settings. The most popular settings are public hospital (30%, n=126), technology or health IT companies (19%, n=79), and state government (14%, n=60). While this is to be expected due to the significant focus on EMRs within hospitals, we need to examine the applicability of the HI framework and CHIA exam to those in primary and community care, and also to raise awareness amongst this healthcare segment.

CHIA candidates have a range of job roles, the most popular roles are health IT professional (28%, n=118), clinicians, including physicians, nurses, and allied health (20%, n=84), and health information manager (10%, n=43). The majority of CHIA candidates have a clinical educational background (23%). Others include business (10%), IT (10%) and health information management (6%).

In 2017, we are embarking on a survey of all CHIA registrants to gather quantitative data. However, the anecdotal qualitative data we have, indicates that the most common reason for people failing or choosing to forfeit is a lack of adequate preparation for the exam. These people underestimated the amount of time required to prepare. To address this, we are modifying our communication with candidates to emphasize that successful attainment of CHIA requires a thorough knowledge base, which requires study, regardless of the amount of work experience the candidate has. We are also making available the reading materials/study guide for those who wish to purchase it ahead of registering for the exam (at present the reading materials are only made available after registration).

Conclusion

CHIA is a unique credentialing program for health informatics. The CHIA credential demonstrates that candidates meet the [Health Informatics Core Competencies](#) to perform safely and effectively as a health informatics professional. Following the early success of CHIA, several Australian universities are currently evaluating how they can include the Australian HI competencies in their training programs, and help prepare graduates to sit to the exam.

The development of this competency framework is helping to raise the profile of health informaticians, contributing to a wider recognition of the profession, and defining more clearly the body of knowledge underpinning this discipline. This framework can also be used as a set of guidelines for recruiting purposes, definitions of career pathways, or the design of educational and training activities. The framework is being re-evaluated in late 2017, and will be re-evaluated every 3 years.

Notes:

¹ Degree can be any degree and does not have to be in health informatics.

² Associated experience includes the following functional areas: health informatics, health administration/management, clinical information systems, e-health, information systems, health information management.

³ <http://www.healthinformaticscertification.com/products/>

Acknowledgments

We would like to acknowledge the below individuals who generously gave their time and expertise to CHIA:

COACH (Alison Delle, Neil Gardner, Don Newsham); HISA (Nigel Chartres, Suzanna Zhang); ACHI (Heather Grain, Peter Williams); HIMAA (Vicki Bennett, Ralph La Tella, Richard Lawrence, Sallyanne Wissmann); HINZ (David Parry, Liz Schoff, James Warren; members of the founding CHIA Examination Committee & Project Committee (Frida Cheok, Peter Croll, David Evans, Anneke Fitzgerald, Jo Foster, Kathleen Gray, Terry Hannan, Sonya Hilberts, Anthony Maeder, Geoff Sayer, Robert Webb, Sue Whetton, Trish Williams, John Wilson); current CHIA Board (Travis Ingram, Kerry Butler Anderson, Klaus Veil, Chris Pearce); current examination committee (Jen Bichel-Findlay, James Walters, Mike Bainbridge, Brendan Wickam, Joy Smith, Paul Macdonald, David O'Driscoll, Robin Mann).

Table 2 – Comparative analysis of competencies in Health Informatics

Competencies/Org	AHIEC	IMIA	AMIA	COACH	CHIA
Scientific skills			✓		✓
Health Informatics	✓	✓	✓		✓
ICT	✓	✓	✓	✓	✓
Information Science	✓	✓	✓	✓	✓
Health Science	✓	✓	✓	✓	✓
Management Science				✓	✓
Areas of specialisation	✓	✓			✓
Human and Social Context			✓		✓
Total number of competencies	45	48	22	51	52
Different levels of competency	YES (6) 1: Know 2: Understand 3: Apply 4: Enable 5: Advise 6: Strategise	YES (2) IT user & HBI specialist	NO	NO	YES 1: Knowledge 2: Comprehension 3: Application 4: Analysis
Specialisation	YES: Health and Aged Care HI IT HI IS HI Specialist HI, Clinical Informatician, Health Information Manager, Clinical Terminologist	YES: 8 focus areas (Imaging, bioinformatics, chemoinformatics, public health informatics)	NO	NO	YES: Combining both IMIA and AHIEC specialties

References

- [1] Education directory. ACHI. 2016. Available at: <http://www.achi.org.au/educationdirectory.htm>.
- [2] F. Martin-Sanchez and K. Gray, Recognition of health informatics in Australian standard classifications for research, occupation and education. *Stud Health Technol Inform* **204** (2014), 92-7.
- [3] Adapted from "What is a unit of competency," by HIMAA, 2012. Available at: <http://www.himaa2.org.au/education/?q=node/90>.
- [4] Adapted from "Overview of Competencies & Benefits and Uses of a Competency-Based System," by P.Shannon and S.Deng, 2008. Available at: <http://www.slideshare.net/davidgay/overview-of-competencies-benefits-and-uses-of-a-competencybased-system>.
- [5] C.A. Kulikowski, E.H. Shortliffe, L.M. Currie, P.L. Elkin, L.E. Hunter, T.R. Johnson, I.J. Kalet, L.A. Lenert, M.A. Musen, J.G. Ozbolt, J.W. Smith, P.A. Tarczy-Hornoch, and J.J. Williamson, AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc* **19**(6) (2012), 931-8.
- [6] J. Mantas, E. Ammenwerth, G. Demiris, A. Hasman, R. Haux, W. Hersh, E. Hovenga, K.C. Lun, H. Marin, F. Martin-Sanchez, G. Wright, IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. *Methods Inf Med* **49**(2) (2010), 105-120.
- [7] COACH. (2012, November), Health Informatics Professional Core Competencies. Available at: <https://www.coachorg.com/en/resourcecentre/resources/Health-Informatics-Core-Competencies.pdf>.
- [8] AHIEC (2011), HI Scope Careers and Competencies Version 1.9, November.
- [9] L.W. Anderson, D.R. Krathwohl, P.W. Airasian, K.A. Cruikshank, R.E. Mayer, P.R. Pintrich, J. Rath, and M.C. Wittrock (eds) A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives, 2000.

Address for correspondence

Prof Fernando Martin-Sanchez. Environmental and Participatory Health Informatics (ENaPHI) Research Group, Division of Health Informatics, Department of Healthcare Policy and Research Weill Cornell Medicine, New York, NY, USA.

425 E 61st Street, Suite 301. 10065 New York, NY. USA Email: fem2008@med.cornell.edu

Information and Communication Gaps in Intersectoral Healthcare Processes for Dementia Patients

Bianca Oppermann^a, Bettina Zippel-Schultz^b, Uta Augustin^b,
Andrea Popa^c, Carsten Schultz^c, Reinhold Haux^a

^a Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig - Institute of Technology and Hannover Medical School, Braunschweig, Germany

^b German Foundation for the Chronically Ill, Fürth, Germany

^c Institute for Innovation Research, University of Kiel, Kiel, Germany

Abstract

In addition to general barriers of information exchange and communication in intersectoral healthcare processes, the healthcare process of dementia patients has unique requirements regarding the communication and information exchange of the various healthcare professionals and non-professional caregivers, such as relatives or their legal caregivers. We conduct a process analysis to determine the status quo of such healthcare processes in the region of Darmstadt, Germany and elicit existing gaps in the information and communication exchange. We identify key processes by performing a document analysis and conducting interviews with seven different healthcare professionals. The results of a weak point analysis based on advanced event driven process chains (EPC) point out 32 information and communication gaps summarized in four categories: "information exchange", "information transfer" "information provision" and "activity impulse". Our results show further opportunities regarding the support of caring relatives and primary care physicians.

Keywords:

Dementia; Intersectoral Collaboration; Workflow

Introduction

Dementia is increasingly affecting people worldwide - in many regions, between five and seven percent of persons over the age of 60 have dementia [1]. Oftentimes, dementia patients suffer from multiple other diseases (multi-morbidity). For example, fractures are a commonly reported co-morbidity of dementia [2] due to impairments of cognitive functions related to dementia [3] and the resulting increased fall risk [4]. In order to address these different diseases, dementia patients are often in need of healthcare services from different healthcare providers and institutions, such as hospitals, specialists or care facilities.

Furthermore, because dementia patients are often neither able to provide information regarding their general state of health nor provide specific details in acute situations, the barriers of communication and information exchange existing between various professional and non-professional actors involved in healthcare provision, are exacerbated in this setting. This is particularly prevalent in unfamiliar situations, like a hospitalization following a fall or other traumatic experience. Hence, in order to offer optimal patient-centered care, healthcare providers involved should coordinate their services (intersectoral interlinkage). However, because the German Social Security System is divided into two separately

organized sectors – inpatient and outpatient care – communicating and cooperating (i.e., sharing information) across these sectoral boundaries is difficult [5]. New care concepts are needed to address these barriers and improve healthcare provision - strategic digital transformation may be instrumental in overcoming these challenges.

The collaborative project "Securing integrated care for multimorbid patients with dementia using an IT-based service concept" (SimPat) aims to (1) identify key processes (status quo) and deduce information and communication gaps in the care process of dementia patients following a fall experience to solve such problems, (2) determine caregivers needs, (3) develop and implement a resulting IT-supported case management system, and (4) evaluate the effectiveness of the IT-solution.

Even though many projects begin with a number of process analyses, the unique situations surrounding the care of dementia patients such as the limited patient participation and the complex, intersectoral care architecture, and the heightened social relevance of dementia as a healthcare challenge, we think it is important to publish the results of our process analysis at this early stage in the project. Therefore, we present the methods and results of the process analysis (point 1) in this paper.

Methods

Setting

The process analysis focuses on multi-morbid patients with dementia after a traumatic fall event, e.g., a femur fracture, with subsequent inpatient treatment. The analysis period is defined as the time from hospitalization up to three weeks of post-treatment follow-up. The process data is collected at a local hospital (AGAPLESION ELISABETHENSTIFT Darmstadt) and various post-hospital care facilities in the region of Darmstadt.

Establishment of the status quo of care

Data collection

The data collection took place in two steps. First, we performed a document analysis of hospital-specific documents, geriatric assessments, and required documents from the nursing home. This basic knowledge was used to prepare the subsequent interviews with the different healthcare professionals that are involved in the examined process period.

Second, we conducted semi-structured, open-ended, interactive interviews with various healthcare professionals involved in the target care period. These interviews had an average duration of one hour. One representative (expert) from each category of professional healthcare providers and a self-help organization was interviewed by two researchers during each interviewer, except for the first interview where 2 respondents participated. We conducted a total of six interviews with seven respondents over a period of three months: one outpatient nurse, one hospital-based social worker, one geriatrician, one primary care physician, one nursing home social worker, and one representative from the local Alzheimer's society.

The interviews started with participants reporting their tasks and activities within the care process without interruption from the interviewers. Meanwhile, the interviewers illustrated this process information as a simple flow chart documenting only activities utilizing a flipchart supplement. This enabled participants to retain a continuous overview of the whole process as they were reporting their tasks and activities. This also facilitated an immediate enquiry on single activities for a more precise documentation of the status quo, especially regarding interfaces to other healthcare professionals.

Upon completion of the interviews, we compared the information of the document analysis with the process data generated from the interviews. In doing so, we were able to supplement any lacking entities, organizational units, and documents used and/or updated in the processes elicited from the document analysis.

Identification, modelling and description of processes

The information collected was summarized in an intersectoral care process, integrating the actions of the healthcare providers with those of the non-professional caregivers. Due to the complexity of this (intersectoral) healthcare process, the combined information was modeled as a simplified process flow chart (only activities and responsibilities), and divided into four (main) process-sections:

1. Admission and treatment at the emergency department, including surgery
2. Post-surgery therapy under the leadership of the trauma surgeon
3. Early rehabilitation under the leadership of the geriatrician, discharge management and discharge from hospital
4. Follow-up at home or in a nursing home.

The resulting flow chart provides a general overview of the comprehensive care process and illustrates relevant key processes for each part as well as sub-processes of intersectoral care.

An additional, more detailed visualization was conducted using advanced event driven process chains (EPC) to model the status quo of the key processes as accurately as possible. The model comprises entities, functions, organizational units, supporting systems as well as input and output information, material, or resource objects to conduct the subsequent weak point analysis. We then used those models to conduct the subsequent weak point analysis. Figure 1 exemplifies the depiction of a sub-process, describing the respective key process in more detail.

Event driven process chain-based process analysis

The EPC-based process analysis consists of three steps: (1) the establishment of performance indicators, (2) the analysis of the modelled EPCs and (3) the preparation and verification of the analysis results.

Establishment of performance indicators

First of all, performance indicators for the process need to be established in order to analyze the EPCs. Generally, a (sub-) process can be considered to be inadequate if it is disrupted or delayed. Depending on the aim of the process analysis to determine information and communication gaps in the care process, the influencing factors considered are the communication and the flow of information between the different healthcare professionals. Consequently, the following performance indicators can be used to analyze the current process in order to detect information and communication gaps: (1) information being unavailable, (2) information being not transferred or not exchanged, or (3) information being disclosed in an untimely manner.

Analysis of event-driven process chains

The analysis of the modelled EPCs based on the preassigned performance indicators was performed by three researchers (researcher triangulation). First, all gaps from the EPCs with regards to unavailable, not transferred or exchanged, or untimely disclosed information were extracted. After that, the identified weak points were aggregated and assigned to different problem areas - categories. The definition of categories was performed on the basis of an inductive derivation using the interviews and the overall process representation (open-ended coding). Figure 1 exemplifies the implemented method, detailing the modeled sub-process "initial anamnesis" within the key process "admission and treatment at the emergency department". The analysis of this EPC-detail highlights the lack of information from the primary care physician at patients' admission to the hospital.

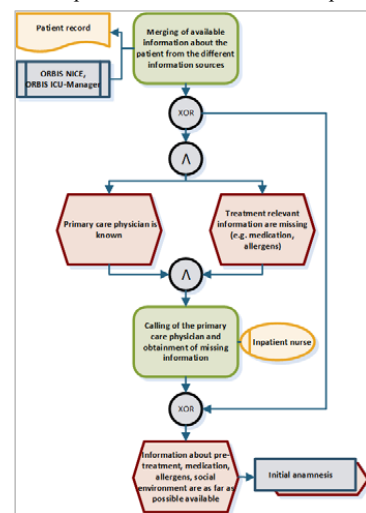


Figure 1– Detail of the sub-process “initial anamnesis” at the emergency department

Preparation and verification of the analysis results

The preparation of analysis results implies, amongst other things, a detailed description including reasons and consequences for each identified gap as well as the accompanying process part on which the gap could be identified. Furthermore, the healthcare professionals involved and the related direction of information flow (uni-, bi- or multidirectional) are documented for the assignment to a problem area. Following, the identified and prepared gaps were presented and verified in the project consortium (member checking). This interdisciplinary team of experts is

composed of medical informaticist, economist and different healthcare professionals from the hospital.

Results

Status quo of care

Overall, we identified five key processes, one for each process-section 1 to 3. Process-Section 4, the follow-up, was divided into two processes depending on the follow-up: (1) follow-up at home, or (2) follow-up at a nursing home. Furthermore 16 related sub-processes, as shown in Table 1, were determined.

Table 1– Identified key processes and related sub-processes

Key processes	Sub-processes
1 Admission and treatment at the emergency department	1.1 Initial anamnesis
2 Treatment at the surgical/geriatric ward	2.1 Admission at the ward and surgical treatment
	2.2 Treatment under the leadership of trauma surgeon
	2.3 Inpatient early rehabilitation
	2.4 Relatives initiated contact to the primary care physician
3 Social counselling and discharge planning	3.1 Planning and organization of aftercare at home
	3.2 Planning and organization of aftercare at a nursing home
	3.3 Clarification of admission related questions
	3.4 Social workers initiated contact to the primary care physician
	3.5 Verification and replenishment of social anamnesis
4 Treatment at home	4.1 Support by an outpatient nurse
	4.2 Care at day-care hospital
5 Treatment at nursing home	5.1 Admission to nursing home
	5.2 Treatment at nursing home
	5.3 Prescription of further measures
	5.4 Planning of an elective surgery

Key process 1 - The care process of dementia patients following a fall event starts with the admission and initial treatment at a hospital, or more precisely, in the emergency department. Depending on the previous care of the patient (treatment at home or in a nursing home) and the type of admission (admission through the primary care physician or as an emergency), varying amounts of information in terms of medication, medical history, and social environment is available for the initial anamnesis (see Table 1 - 1.1). Currently, a short dementia-check takes place in the emergency department to determine, whether the patient needs to have an interdisciplinary treatment due to a manifested dementia. If the patient displays symptoms consistent with dementia, the geriatrician is consulted for a treatment and a risk assessment in the emergency department. Depending on the results of the risk assessment, the patient either undergoes immediate surgical therapy or a conservative, pre-surgical acute geriatric therapy followed by surgery.

Key process 2 - Following the surgery, the patient is transferred to the surgical/geriatric ward for seven days of post-surgery therapy under the care of the trauma surgeon (see Table 1 - 2.1 to 2.2). This is then followed by the inpatient early rehabilitation including nursing and medical treatment as well as physio- and ergotherapy aimed at achieving the best-possible recovery of independence and mobility (see Table 1 - 2.3).

Key process 3 - A continuous social anamnesis is performed by a social worker in the surgical/geriatric ward. Its purpose is to adjust the individual treatment to the patients' needs and to timely plan discharge (see Table 1 - 3.1 to 3.3). Here, information about the patient (e.g., social environment, demands and characteristics) are continuously collected. The collection of information starts with the admission of the patient at the ward. Different sources (e.g., relatives, neighbors, present nurses) are used to gather the necessary information. At the latest, the social anamnesis must be completed before the concrete planning of discharge time and aftercare possibilities starts (see Table 1 - 3.5). To guarantee an adequate and smooth planning of discharge and aftercare at home or in a nursing home, communication and information exchange between relatives and hospital, hospital and follow-up care givers (i.e., outpatient nurse or nursing home) as well as relatives and follow-up care givers is essential (see Table 1 - 3.1 to 3.2). In addition, it is very important to integrate the primary care physician (if known) as soon as possible (see Table 1 - 3.4).

Key process 4 and 5 - Depending on the patient's situation before admission, the social environment and the condition of the patient, the patient can either be discharged to his or her home or to a nursing home. Should the patient be discharged to a nursing home, care-relevant questions concerning the care level, financing, or necessary health aids, must be settled before admission can take place (see Table 1 - 5.1 to 5.2). Should the patient go home, relatives or legal caregivers must determine whether the support of outpatient hospital services or an in-home nurse is necessary (see Table 1 - 4.1 to 4.2). Regardless of where the aftercare takes place, the primary care physician is responsible for the subsequent therapy, including the prescription of medication and any necessary health aids as well as the suggestion and planning of an elective surgery (see Table 1 - 5.3 to 5.4).

Definition of gap categories

To classify the identified weak points – gaps – into various problem areas, we defined the categories shown in Table 2. Here, the designation does not refer to a possible solution but rather to the existing problem.

Table 2– Gap classification and definition

Designation	Definition
Information transfer	Active disclosure of information from one healthcare partner to another (unidirectional).
Information provision	Passive disclosure, therefore provision of information for one or more healthcare partners.
Information exchange	Mutual disclosure of information between two healthcare actors (bi-, or multidirectional).
Activity impulse	Encouragement of an activity which delays the executive process when not or delayed performed.

Identified gaps

In sum, we identify a total of 32 weak points within the four defined categories: 6 in “information transfer”, 13 in “information provision”, 4 in “information exchange” and 8 in “activity impulse”. Tables 3 to 6 show the identified gaps according to the pre-defined categories (Table 2).

Table 3– Gaps concerning “information transfer”

No.	Lack of information transfer...
1.1	... to primary care physician about hospital admission
1.2	... to primary care physician just before hospital stay
1.3	... to hospital about existing care at home
1.4	... to outpatient nurse about aftercare demands and health aids
1.5	... to outpatient nurse about hospital admission
1.6	... to primary care physician about patients death

The results show a particular lack of information transfer (Table 3) and information provision (Table 4) from the hospital to the primary care physician during the hospital stay. Commonly, the primary care physician is neither informed of a patients’ admission and discharge nor about the treatments undertaken. The primary care physician only receives this information if the hospital has treatment-relevant questions at admission (and the primary care physician is known) or if relatives inform the primary care physician about the admission.

Table 4– Gaps concerning “information provision”

No.	Lack of information provision...
2.1	... from local Alzheimer society
2.2	... about the primary care physician at hospital admission
2.3	... about the patient’s condition after discharge
2.4	... about contacts for specific questions to caring relatives
2.5	... about prescribed drugs
2.6	... about changes of prescriptions
2.7	... about inpatient physio- and ergo therapy (within the discharge documents)
2.8	... about outpatient care possibilities
2.9	... about the treatment at hospital to the primary care physician
2.10	... about nursing relevant questions to relatives
2.11	... from the outpatient nurse to the hospital at admission
2.12	... about the patient to the outpatient nurse before discharge
2.13	... about existing medical and nursing care of the patient to primary care physician

The results show that the healthcare providers have very little to no joint knowledge about the patient. Information about their medication regimens, social environment, and current treatment is often missing or incomplete. This is usually not an issue as the patient is often able to fill these gaps by transmitting missing information themselves, e.g., to the primary care physician. However, as we describe above, patients with dementia may be highly restricted in their communication and cognitive abilities and as a result, cannot provide the necessary information. For instance, should the hospital change the current treatment and medication plan, this information along with the driving reasons behind the change is lost, so that ultimately, the primary care physician has no

knowledge of the change and therefore cannot uphold the new regime, reverting the treatment back to the original regimen. The primary care physician also has difficulty accessing this information. The categories “information provision” (Table 4) and “information exchange” (Table 5) address those problems.

Table 5– Gaps concerning “information exchange”

No.	Lack of information exchange...
3.1	... between the different healthcare professionals during the anamnesis at hospital
3.2	... between outpatient nurse and hospital for discharge planning
3.3	... between hospital and primary care physician about previous medical treatment
3.4	... between the different healthcare professionals concerning the medication plan (joint access)

Occasionally, some parts of the healthcare process are interrupted or delayed by the failure to complete preceding activities on time. This is not only attributable to a lack of information exchange and communication, but rather to a missing impulse that must be given automatically or from one actor to another to initiate the activity. As seen in Table 6, caregiving relatives are affected in a number of ways. During the hospitalization, they must complete many tasks in order to ensure an appropriate follow-up at home or at nursing home. However, they are often not aware of when and where information can be accessed, or at which time certain information needs to be transmitted. Especially, tasks such as 4.3, 4.5 and 4.8 are neglected by the caring relatives due to the growing work load as well as the physiological burden.

Table 6– Gaps concerning “activity impulse”

No.	Missing impulse...
4.1	... to perform the detailed social anamnesis
4.2	... to initiate necessary actions of non-professional caregivers regarding medical aids and appliances prescriptions and application
4.3	... to bring discharge folder to re-admission of patients
4.4	... to perform dementia-check in the hospital at the right time
4.5	... to inform primary care physician about discharge of hospital by the non-professional caregiver
4.6	... to sign the discharge documents by two physicians
4.7	... to inform outpatient nurse at hospital admission
4.8	... to clarify questions for the admission at nursing home by the non-professional caregivers

Discussion

This process analysis attempts to describe the status quo of the intersectoral healthcare processes of dementia patients after a traumatic event in the region of Darmstadt. To achieve this objective, we determined the intersectoral care processes and identified information and communication gaps between healthcare professionals and non-professional caregivers. Using EPCs and an EPC-based weak point analysis, we illustrate the identified key and sub-processes, and classified information and communication gaps. These modelling methods enable a simplified depiction of complex processes with parallel and alternative activities [6]. Furthermore EPCs can be used “[...] to find parts of the process that are never

used or find parts where users deviate from the prescribed procedure" [7], and are thus instrumental in also identifying weak points or gaps.

The results show that significant gaps exist in the process concerning the transfer, provision, and exchange of information. The gaps often occur when coordination or communication between the healthcare sectors is necessary for a seamless process, such as the transfer and provision of information between a hospital and a primary care physician. Moreover, contrary to healthcare professionals, non-professional caregivers are often unfamiliar with typical activities in the care process, and are unaware of the activities expected of them. For example, when a loved one suffering from dementia is admitted to the hospital due to a fall, caregiving relatives may be unaware that the hospital needs basic information such as patient history and a medication plan in order to optimally plan further treatment steps. Reminders or information requests, or other activity impulses, may or may not be in place to collect this information from external parties. Impulses, given by the healthcare professionals, may in fact initiate those activities. Our analysis lead us to deduce that the care process may be disturbed when those impulses are missing.

There are, however, some limitations to our study. In order to identify the key processes, we conducted an extensive document analysis. The document analysis refers to the documents of one hospital in the region of Darmstadt. The hospital, previous to our study, established a structure promoting a tight cooperation between the surgery and geriatric ward, placing special attention to the care of patients with dementia. During the interviews, some respondents emphasized the fact that the communication with this hospital is better than the communication with other hospitals in the region. On the one hand, this must be considered when interpreting the results (transferability). On the other hand, we were still able to identify gaps that represent general requirements for a functioning intersectoral care process. We also interviewed a limited number of healthcare professionals. However, although the perception of the care process might differ between individuals, the activities within the process remain similar. The key processes were described in a very similar manner and respondents also supplemented the key processes reported by defining the sub-processes.

Finally, limitations of the method implemented are manifest in the modeling of processes. Due to the limited number of modelling components and modelling rules, an intersectoral illustration with a large number of interfaces, overlappings and merged processes is very complex and hardly executable.

Conclusion

In order to provide optimal patient-centered care, healthcare providers involved in the care process should coordinate their services (intersectoral interlinkage). However, because the German Social Security System is divided into two separately organized sectors – inpatient and outpatient care – communicating and cooperating across these sectoral boundaries is difficult. Dementia patients are especially in need of integrated care, because they are often neither able to provide information regarding their general state of health nor provide specific details in acute situations. Because of the unusual demands caring for dementia patients in acute medical situations places on professional and non-professional actors, as well as on the patients themselves, we publish the results of our process analysis to further the field and encourage discussion on important developments in this field.

In this paper, we analyze the intersectoral care process of multimorbid dementia patients that have suffered a traumatic event with a subsequent hospital admission to identify key processes and weaknesses in the care process. We determine the existence of five key processes in the care process and identify and describe 32 gaps in the information and communication exchange of the care process that have the potential to significantly impact the quality and efficiency of care provision. We also find that healthcare professionals may be in a unique position to generate activity impulses in order to overcome these communication and information deficiencies, thereby securing optimal care for the patients.

To solve such identified problems, further steps involve the establishment of functional and non-functional requirements as well as the development and implementation of an IT-supported case management system supporting the communication and information exchange beyond the sectoral boundaries via a holistic view at the patient.

Acknowledgements

This work is part of the collaborative research project SimPat funded by the Federal Ministry of Education and Research, funding code 02K14A024.

The authors would like to thank all project partners for their assistance, especially AGAPLESION gAG (Claudia Möller) and AGAPLESION ELISABETHENSTIFT Darmstadt (Andrea Maetzel) for supporting the recruitment of interview-partners. Special thanks is due to the different interviewed experts who took the time to answer all our questions with patience and endurance.

References

- [1] Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri C.P. The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimers Dement* 9 (2013), 63-75.
- [2] Sanderson M, Wang J, Davis D.R, Lane M.J, Cornman C.B, Fadden M.K. Co-morbidity associated with dementia. *Am J Alzheimers Dis Other Dement* 17 (2002), 73-78.
- [3] World Health Organization, Alzheimer's Disease International, Eds. *Dementia: a public health priority*. Geneva: World Health Organization; 2012.
- [4] J.C. Davis, J. Best, C.L. Hsu, L.S. Nagamatsu, E. Dao, and T. Liu-Ambrose, Examining the Effect of the Relationship Between Falls and Mild Cognitive Impairment on Mobility and Executive Functions in Community-Dwelling Older Adults, *Journal of the American Geriatrics Society* 63 (2015), 590-593.
- [5] Busse R, Blümel M. Germany: health system review. *Health Systems in Transition* 16 (2014):1-296.
- [6] M. Dumas, W.M. Van der Aalst, and A.H. Ter Hofstede, *Process-aware information systems: bridging people and software through process technology*, John Wiley & Sons, 2005.
- [7] J. Recker, J. Mendling, W. Van Der Aalst, and M. Rosemann, Model-driven enterprise systems configuration, in: *International Conference on Advanced Information Systems Engineering*, Springer, 2006, pp. 369-383.

Address for correspondence

Bianca Oppermann, M.Sc.
 Peter L. Reichertz Institut for Medical Informatics
 Mühlenpfordstr. 23
 38106 Braunschweig Germany
 eMail: bianca.oppermann@plri.de

Equality Challenges in the Use of eHealth: Selected Results from a Danish Citizens Survey

Lone Stub Petersen, Pernille Bertelsen

Department of Development and Planning, Aalborg University, Aalborg, Denmark

Abstract

The increased focus on and use of citizen/patient generated health data has spurred a wide range of personal health technology projects within digital intervention in health, e.g. telehealth. These developments are focused on objectives of improving health, but also claiming to provide health services at a lower cost. However, the question is: do the ways healthcare technologies are designed and developed support and improve healthcare services for those who need it the most?

Survey data from our study point toward health informatics challenges in reaching the people who are considered in 'health risk' group, who, in this study, are interpreted as people with low level of education. The study shows that this group is less likely to use and communicate through health information technologies (HIT) and is generally more sceptical towards the benefits of HIT.

We conclude that there is a need to pay specific attention to the patient groups that are socio-economically and health wise weakest during HIT design and development. It would also provide equality and equity in digital health intervention and access to healthcare for them in the future.

Keywords:

Surveys and Questionnaires; Educational Status; Health Services

Introduction

Today we see an increased focus on the use of patient-generated health data (PGHD) both developed for consumers interested in improving their own health, and for use in the healthcare sector [1], where many healthcare professionals and health care managers see an enormous potential in HIT to improve service and reduce costs. The potential usefulness is seen as enormous but we know little about a) who uses the technologies and b) how these data can benefit/support health care services [2].

Table 1 – Two categories of data generated by patients

Citizens - person/patient-generated health data (C-PGHD)	Healthcare professional - person/patient-generated health data (HP-PGHD)
Generated by citizens/patients for citizens/patients	Generated by citizens/patients for healthcare professionals
Controlled and owned by the citizens/patient	Controlled and owned by the healthcare professional

The two perspectives on PGHD presented in Table 1 have data generated by patients as common ground but differ when it comes to the objective or purpose of why patient data is collected and also when it comes to purpose of its use. The commonly used terms, when it comes to framing the intersection between citizen or consumer generated data and the use or non-use of these data by health care professionals are patient or person generated data (PGD) and/or patient generated health data (PGHD). Person or patient generated health data (PGHD) can be defined as health-related data created, recorded, or gathered by patients (or by their family members or other care givers) for own use or to be shared with their health care providers [3]. We will go into depth first with the citizen/patient perspective and then the health care professional perspective in the following section.

Citizens perspective in HIT and patient generated health data. (C-PGHD)

Citizens and patients are beginning to use a range of health information technologies (HIT) to track and generate personal health data outside the clinic. Today, numerous types of HIT and more contemporary consumer health technology are migrating from being expensive "first movers" toys to becoming inexpensive and generally available to the more average consumers. These technologies provide patients the ability to track their own health data [1]. Some of the best known platforms supporting and encouraging people with smartphones to track their own health are Apple's HealthKit [4] and Google's Google Fit [5]. The potential of using PGHD data are seen as enormous but it still remains to be seen how best to integrate these data into contemporary clinical work, and to make sure those in most need also are those being encouraged to collect own health data and thus benefit from the use of the health data.

Market developments predict that consumer technology for health management will continue to increase – it has been estimated that wearable health technology will become the eighth largest revenue driver within consumer and mobile devices in 2018 [6]. This increase shows strong indications on citizens' interest in everyday health management. Simultaneously, this has sparked increased interest from health professionals to explore how to include patient generated health data as part of clinical practice and electronic health records [3].

This development, often framed as Quantified self or life logging, enables patients to capture, measure, track, and analyse data from and within their daily life (e.g. physical activity, food consumption, mood). Availability of relatively inexpensive wearable technologies and availability of digital storage of personal data offline and on-line has made personal HIT a research topic [7]. Now citizens are able to gather and store

large volumes of personal data in a very cheap manner, e.g. by using their smartphones.

This availability of data put pressure on the health professionals to include these technologies in their treatment. How the challenge materializes in the clinical setting depends, among other things, strongly on the type of national health care system governance, for example, the degree of public, private, and health insurance financed health interventions. The healthcare sector will need to meet the challenge by developing and implementing digital health interventions (DHI) [8] and by enrolling patients.

Healthcare Professionals, eHealth and patient generated health data (HP-PGHD).

Citizens and patients are still in the upstart of exploiting HIT to either support their own management of health and diseases at home or to supplement the service provided by their health care provider [8]. DHI aimed at the public or citizens in general have a wide span from telehealth and telecare systems to mobile health applications and devices [9, 10]. As a result, new HIT tools are becoming available for clinical practises to potentially be integrated in their health interventions, and are also available at the market for citizens to buy and use in their private sphere. Mobile technologies for self-monitoring of wide-ranging variables such as blood pressure, physical activity, blood glucose, daily weights, sleep etc. are available and question the traditional patient-physician relationship [11].

Challenges of PGHD when used in the clinical context of healthcare professionals include the reliability and accuracy of the data collected, forgetfulness of patients, e-health literacy, attitudes towards technology, and patients' self-bias [12]. However, the innovative technologies and the change of culture related to Quantified Self have opened up the potential for patients to collect more accurate and reliable health data [3].

In addition to DHI, another form of PGHD used in clinical settings are Patient reported outcomes (PRO) and Patient Reported Outcome measures (PROM). PRO is information on patients' health and quality of life, reported directly from the patient to be used by health provider organizations as a feedback and quality assurance measurement system on local as well as national levels. PRO is information about a patient's health, including physical and mental health, symptoms, health related quality of life, and functional level [12, 13, 14]. Patient reported data could inform different aspects of patient experience, from their encounter with the healthcare sector, for example, the structure, processes and results [16]. Patient Reported Outcome measures (PROM) may be seen as a subcategory of PRO where the systematic collection of data on patient experiences are measured at, for example a national level, as is seen with NHS in England [13], and the Danish Cancer Association [14].

PRO and PGHD are both concepts that feed into a debate on who uses and how to make use of and learn from, the data patients generate in different health care provider settings and while developing DHI. PRO and PROM, as well as tele health information are structured by and feed into clinical practice in a format that clinicians have control over. This is not the case with C-PGHD. These data are not formally structured but have the same quality assurance issues as the HP-PGHD. This makes it difficult for clinicians to agree on their usefulness in daily clinical practice and therefore to integrate these data into the existing healthcare services.

Health inequality – The Socio-technical challenge of HIT

In this context of C- and HP-PGHD, we like to argue for the importance of a social investigation of who the users of the HIT are, and when doing so keep an eye on whether the in-

creased use of HIT in healthcare is decreasing or increasing health inequality in society. The World Health Organisation (WHO) defines health inequalities as “avoidable inequalities in health between groups of people within countries and between countries” and describes health inequality as significantly influenced by social determinants [15]. It is paradoxical that despite a long tradition as a welfare country, and a relatively low income gap between rich and poor, Denmark is a country that within the last 25 years has doubled health inequality (measured on mortality) [16]. This inequality does not only need to be handled through healthcare governance, but needs to be addressed when developing HIT.

Showell and Turner from an Australian context [17] argue that most HIT claiming to be of use for patients, are designed and developed, and therefore being used, by “people like us” (PLU). ‘People like us’; that is, “people who believe to understand healthcare and health issues, take care of their own health, are literate, well to do, tech-savvy, and hold a tertiary qualification” [17]. On the other hand, those citizens who really are in strong need of better health, care, and support are the disempowered, disengaged and disconnected, the DDD's. If this is indeed the case and most health promoting IT used in DHI reinforces health inequality, there should be a motivation for understanding and contributing to change in development of HIT.

In this paper we report on selected findings from the 2015 survey based on data from eight questions (20A, 20I, 20B, 22A, 21A, 23, 24,) [18] that explore the citizens' experiences with accessing and using a) the national health portal Sundhed.dk, b) consumer apps and internet and c) mobile health apps and services. Educational background data have been analysed to investigate if a difference in use of IT correlate with respondent's educational background. Education is free for all citizens in Denmark, and has been so for more than a century. Therefore, all citizens have at least a primary school education. Further, educational level, socio-economic status, and health are strongly related in a Danish context. This is the reason education is used as an indicator of socio-economic and health status in this analysis of inequality. We acknowledge that this might not be the case in all countries, however, we are confident that similar challenges of the reproduction of inequality in eHealth can be found in other countries.

Methods

The survey reported from in this paper is the second in a bi-annual series of National surveys on Danish citizens' expectations and perspectives on eHealth. The first was done in 2013, inspired by Canadian and Australian studies of consumer experience with eHealth [18,19]. The second survey in 2015 was further inspired by questions posed in national surveys from Norway and Finland. The survey is supported by Danish eHealth Observatory and The Danish Center for Health Informatics who have monitored eHealth implementation in Denmark for many years such as the national implementation of the Electronic Health Record (EHR) and the national monitoring of clinicians' use of health informatics in their daily practices. A Danish market research agency (Megafon) was commissioned to carry out the surveys with a population sample of n=1,059 in 2015 and n=1,058 in 2013. The questionnaires were tested twice. The surveys are combination of using both email and telephone. The selected respondents are part of a citizens' panel reflecting the Danish adult population with respect to age, education, and geographic distribution.

Results

People with only primary school education are less likely to use the Danish national health portal than those with higher education (Sundhed.dk in Figure 1). Only 21% of people with only primary school education have used the portal while 60% of people with a high education have used it.

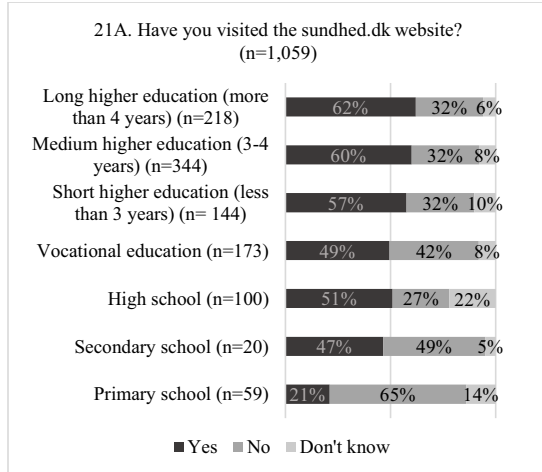


Figure 1 - Cross Table: Educational Level and Expected Impact of Health IT

Citizens with only primary school education have less experience with using applications developed for health purposes (8% vs 26% in Figure 2).

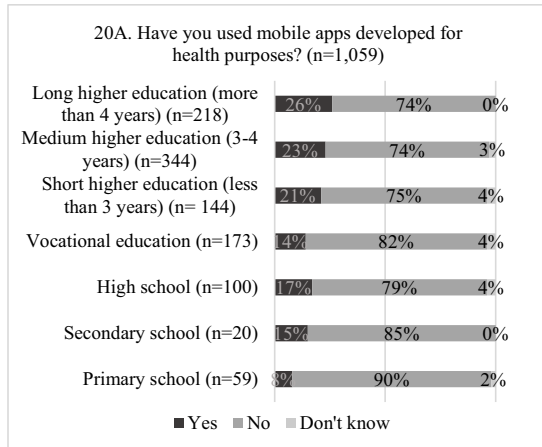


Figure 2 - Cross Table: Educational Level and Expected Impact of Health IT

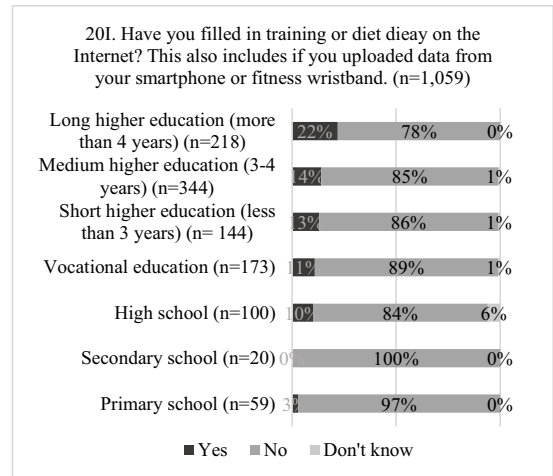


Figure 3 - Cross Table: Educational Level and Expected Impact of Health IT

Few citizens used Internet or fitness trackers to log information about training and food (30% n=1,059 in Figure 3). Also, here citizen with high education are more likely (22%) to use these services than people with low or only primary school education (3%).

The same is the case when asked about the use of Internet based self-help health services (6% low vs. 25% higher education in Figure 4).

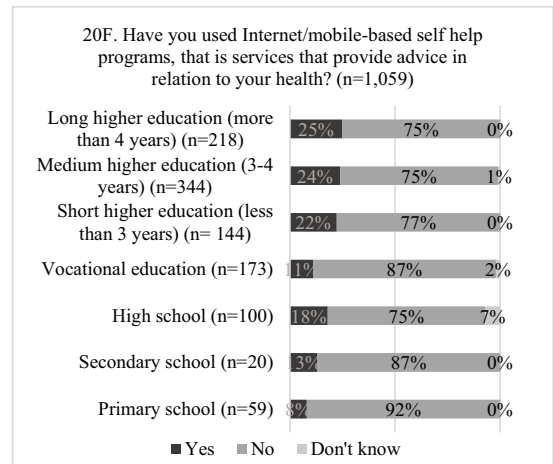


Figure 4 - Cross Table: Educational Level and Expected Impact of Health IT

The citizens with low education are less likely to ask questions about health and symptoms to healthcare professionals on the Internet (14% vs 23% in Figure 5).

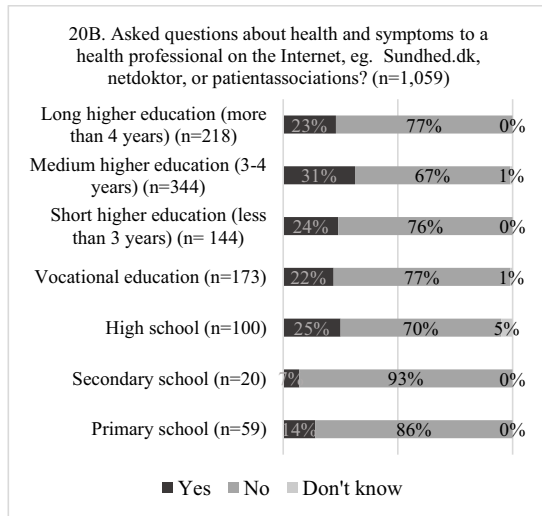


Figure 5 - Cross Table: Educational Level and Expected Impact of Health IT

When asked if they generally agree or disagree to the statement "I have difficulties navigating within health IT systems", 26% with only primary school education agreed while only 14% of people with high education agreed (Figure 6).

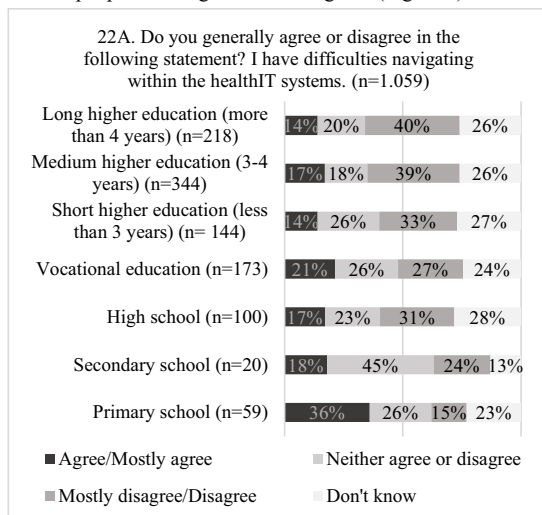


Figure 6 - Cross Table: Educational Level and Expected Impact of Health IT

Danish citizens are generally positive towards the use of IT in healthcare. However, the study shows that the higher the educational level, the more positive they are. As shown in Figure 7, there is high scepticism among respondents with low education when questioned whether they expected health IT to improve or impair the quality of the healthcare services they expect to receive within the next three years.

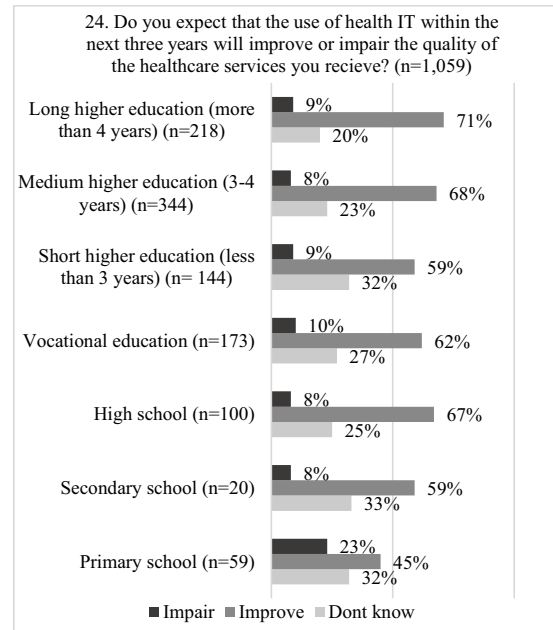


Figure 7 - Cross Table: Educational Level and Expected Impact of Health IT

Discussion

The study shows a significant difference between citizens with low and high educational level, when it comes to opinion on, knowledge of, and experience with the use of HIT and other Internet, apps or mobile services. In the answers, there was a significant tendency that people with low educational background are more sceptical, not so familiar with HIT and find it difficult to use the technologies and systems.

This indicates a strong need to target and support the socio-economically weak population in getting access to technology that can assist in accessing HIT and enrolling in DHI. The findings of the study empirically support the concern and conceptual framing of DDDs and PLU of Showell and Turner [17]. To address the findings, concerns as well as the challenges of health inequality in Denmark, attention needs to be placed on the role of the DDDs in HIT design, development, and use.

DHIs need to be specifically targeted at those citizens with the greatest needs if the digital and health divide is not to grow even larger. If the socio-economically challenged DDDs are to be given access to live a healthier life, it may have a positive impact on the health outcome and thus provide an economic benefit to the health care system.

To summarize this view, we propose (Table 2), that the problem be addressed both in relation to HIT developed to provide citizens with health data (C-PGHD) and DHI providing healthcare professionals with citizen and patient data (HP-PGHD).

Table 2 – Two Categories of Data Generated by Patient – and How to Handle HIT Inequality

Citizens - person/patient-generated health data (C-PGHD)	Healthcare professional - person/patient-generated health data (HP-PGHD)
Generated by citizens/patients for citizens/patients	Generated by citizens/patients for healthcare

Controlled and owned by the citizens/patient	professionals Controlled and owned by the healthcare professional
How to address health inequality in relation to HIT	
Provide special support to engage and involve socio-economically disadvantaged patients in design and use of HIT	Involve socio-economically disadvantaged users in Design (or re-design) of DHI-technology

The HIT technologies used to support such change need to be understood as not only technical but also socio-technical and therefore, designed and targeted with attention to the context of use and their future users. To point at what can be done to increase the equality, user centered design (where designers investigate what users need) [20] is believed not to be an adequate or sufficient approach to overcome the dominating PLU problem. Instead, we—knowing that this does not follow directly from the data in this paper—suggest technology designers to use a User Innovation Approach (UIM) [21]. Here users are motivated and encouraged by facilitators to develop design concepts that meet their specific needs (e.g. the DDDs specific needs). The designers of HIT and DHI need to hand over the stick to the socio-economically disadvantaged people (the DDDs) and by doing so for some time reduce their own professional role to (1) plan and facilitate, and (2) draft design concepts and draft prototypes to be validated and changed by the participants [22, 23, 6].

Conclusion

On the basis of an examination of the educational inequality detected in perspective on and use of HIT in a national Danish survey, we propose an increased focus on inequality in the use and development of HIT and involvement of citizens and patients from the socio-economically disadvantaged groups in HIT design and development. It will be of benefit to technology developers and the healthcare providers to improve the use of appropriate consumer HIT for this group, in both health care services and preventive health initiatives as well as general health monitoring and management in everyday life. Engaging with and involving this group of citizens and patients is important in levelling out inequality in health which is a key concern in a Danish healthcare context.

Acknowledgements

We are appreciative of the Danish Center for Health informatics (DaCHI) for making the survey possible.

References

- [1] M. Almalki, K. Gray, F. Sanchez, M. Swan, M. Almalki, F. Martin-Sanchez, K. Gray, M. Almalki, F. Martin-Sanchez, K. Gray, D. Lupton, S. Fox, M. Duggan, M. Swan, M. Beyer, D. Laney, J. Hibbard, J. Stockard, E. Mahoney, M. Tusler, F.M. Sanchez, K. Gray, R. Bellazzi, G. Lopez-Campos, S. Vashist, N. Gupta, S. Jilla, J. Choi, B. Choi, J. Seo, R. Sohn, M. Ryu, W. Yi, K. Park, T. Wilson, K. Kuutti, I. Li, A. Dey, J. Forlizzi, W. Jones, I. Li, A. Dey, and J. Forlizzi, The use of self-quantification systems for personal health information: big data management activities and prospects, *Heal. Inf. Sci. Syst.* **3** (2015)
- [2] J.G. Anderson, M.R. Rainey, and G. Eysenbach, The Impact of CyberHealthcare on the Physician – Patient Relationship, *J. Med. Syst.* Vol. 27, No. 1, (2003) 67–84.
- [3] H. Zhu, J. Colgan, M. Reddy, and E.K. Choe, Sharing Patient-Generated Data in Clinical Practices: An Interview Study, in: AMIA Annual Symp. Proc., Chicago (2016): p. 10.
- [4] Apple, Apple HealthKit, <https://developer.apple.com/healthkit/>, (accessed April 14, 2017).

- [5] Google, Google Fit, <https://www.google.com/fit/>, (accessed April 14, 2017).
- [6] A.M. Kanstrup, and P. Bertelsen, Bringing new voices to design of exercise technology, in: Proc. 14th Particip. Des. Conf. Full Pap. - PDC '16, ACM Press, New York, New York, USA, (2016): pp. 121–130.
- [7] L. Zhou, and C. Gurrin, A survey on life logging data capturing, *SenseCam*. (2012).
- [8] S. O'Connor, P. Hanlon, C.A. O'Donnell, S. Garcia, J. Glanville, and F.S. Mair, Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies, *BMC Med. Inform. Decis. Mak.* **16**. (2016).
- [9] S.A. Moorhead, D.E. Hazlett, L. Harrison, J.K. Carroll, A. Irwin, and C. Hoving, A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication., *J. Med. Internet Res.* **15** (2013) e85.
- [10] S.C. Bailey, L.T. Belter, A.U. Pandit, D.M. Carpenter, E. Carlos, and M.S. Wolf, The availability, functionality, and quality of mobile applications supporting medication self-management., *J. Am. Med. Inform. Assoc.* **21** (2014) 542–546.
- [11] D.L. Boeldt, N.E. Wineinger, J. Waalen, S. Gollamudi, A. Grossberg, S.R. Steinhubl, A. McCollister-Slipp, M.A. Rogers, C. Silvers, and E.J. Topol, How consumers and physicians view new medical technology: Comparative survey, *J. Med. Internet Res.* **17** (2015) e215.
- [12] K. O'Leary, L. Vizer, J. Eschler, J. Ralston, and W. Pratt, Understanding patients' health and technology attitudes for tailoring self-management interventions., *AMIA ... Annu. Symp. Proceedings. AMIA Symp.* (2015) 991–1000.
- [13] NHS England, National Patient Reported Outcome Measures (PROMs) Programme Consultation - NHS England - Citizen Space. https://www.engage.england.nhs.uk/consultation/proms-programme/consult_view/ (accessed April 14, 2017).
- [14] Kræftens Bekæmpelse, Klinisk anvendelse af Patient Reported Outcome Measures (PROM) – en evalueringsrapport, Copenhagen, 2016.
- [15] C. on S.D. Health, Closing the gap in a generation: health equity through action on the social determinants of health: final report of the commission on social determinants of health, Geneva: World Health Organization, 2008.
- [16] F. Diderichsen, I. Andersen, C. Manuel, A.-M.N. Andersen, E. Bach, M. Baadsgaard, H. Brønnum-Hansen, F.K. Hansen, B. Jeune, T. Jørgensen, and J. Sogaard, Health inequality - determinants and policies., *Scand. J. Public Health.* **40** (2012) 12–105.
- [17] C. Showell, and P. Turner, The PLU problem: are we designing personal health for People Like Us?, Enabling Health and Healthcare Through ICT: Available, Tailored and Closer 183 (2013): 276.
- [18] L.S. Petersen, and P. Bertelsen, Undersøgelse af borgernes perspektiv på sundheds-it i 2015, DaCHI, Aalborg, 2016.
- [19] P. Bertelsen, and K. Tornbjerg, Undersøgelse af borgernes anvendelse af sundheds-it i 2013, DaCHI, Aalborg, 2014.
- [20] A.W. Kushniruk, and C. Nohr, Participatory Design, User Involvement and Health IT Evaluation, in: Evidence-Based Heal. Informatics, 2016.
- [21] A. Kanstrup, and P. Bertelsen, User Innovation Management, Aalborg Universitetsforlag, 2011.
- [22] A.M. Kanstrup, P. Bertelsen, and J.Ø. Madsen, Design with the feet, in: Proc. 13th Particip. Des. Conf. Res. Pap. - PDC '14, ACM Press, New York, New York, USA, (2014): pp. 51–60.
- [23] A. Kanstrup, and P. Bertelsen, Design for more: an ambient perspective on diabetes, Proceedings of the Tenth Anniversary Conference on Participatory Design 2008. Indiana University, (2008). 118–127.

Address for correspondence

Lone Stub Petersen, lonsep@plan.aau.dk.

Analysis of the Updated Swedish Regulatory Framework of the Patient Accessible Electronic Health Record in Relation to Usage Experience

Isabella Scandurra^a, Maria Pettersson^b, Benny Eklund^c, Leif Lyttkens^c

^a Informatics, School of Business, Örebro University, Örebro, Sweden

^b Department of Journalen, Inera AB, Stockholm, Sweden

^c Uppsala County Council, Uppsala, Sweden

Abstract

In Sweden, all citizens can (in 2017) access their health data online from all county councils using one national eHealth service. However, depending on where the patient lives, different information is provided as care providers have assessed differently how to apply the National Regulatory Framework (NRF). The NRF recently was updated and this paper analyses version 2.0 should now serve as the guideline for all county councils. Potential improvements are analyzed in relation to patient experiences of using the service, and the rationale for each change in the NRF is discussed. Two real case quotations are used to illustrate potential implications for the patient when the new version is placed into operation. Results indicate that this NRF allows for opportunities to create a national eHealth service that better supports patient-centered care and improves health information outcome.

Keywords:

Patient Portals; Electronic Health Records; Health Information Exchange

Introduction

The number of eHealth services for patients is rapidly increasing, as many countries are launching these services as a means to manage an ageing population, to increase efficiency in healthcare, and to empower patients. Although aiming to improve patient outcomes and satisfaction [1], the adoption of these services has often not been as successful as anticipated by politicians and vendors [2, 3]. Many healthcare professionals are concerned about how, e.g., the information will be understood by patients when reading without any medical support [4]. Patients, on the other hand, often strive to manage their own health and illnesses and want their electronic health record (EHR) [1, 5, 6]. In short, deployment of such a service is controversial for healthcare professionals and the establishment of an asset of rules regulating how the eHealth service should be used has been much discussed in Sweden, since the first region-wide deployment of the eHealth service enabling patients' access to their EHR in 2012 [7]. Sweden has recently (in 2017) extended the Patient Accessible Electronic Health Record (PAEHR) service to cover all county councils and to include its 10 million citizens.

Also, internationally, similar services have been provided to large groups of patients, such as in the UK, the US, and Norway [2, 6, 8, 9]. The strive towards providing PAEHRs has been limited in part by professional resistance and concerns about security and privacy [10], legal constraints [11] and the

need for interoperable solutions [12]. Little has been studied regarding implementation of policies for PAEHR services [2], although policies form the basis for such services, and their use is ultimately dependent on how they are implemented.

Sweden has a decentralized healthcare system, and a National Regulatory Framework (NRF) was developed to support the deployment of the PAEHR in the 21 county councils. From the point of view of the patient, the first version of the NRF was less successful as it contained electable paragraphs that were applied differently across counties [13]. This resulted in a national eHealth service that displayed health data differently depending on where, when and why a patient sought treatment [13].

This paper analyses the recently updated version (October 15, 2016) of the NRF in relation to the previous one, and the rationale for each change is discussed. The analysis was performed in the light of a service still considered controversial, county councils that act autonomously, and a national development and deployment that are ongoing. As many countries are facing an introduction to eHealth services providing health information to patients, not only the NRF, lessons learned from Sweden may be of interest to policy makers and developers in order to improve deployment and use of PAEHRs and similar eHealth services elsewhere.

Current status of the national PAEHR 'Journalen'

The PAEHR is currently (April 2017) accessible by the citizens of 18 out of 21 healthcare regions or county councils. They access it via a national patient portal www.1177.se, which contains several services and functionalities. Sweden has approximately 10 million inhabitants, where 37.9 % have created their own account for the 1177.se portal in order to reach these services. Statistics of the month of January 2017 show almost 2 million logins. Each user has, per year, logged in 10.7 sessions (~ services) of 1177.se. The PAEHR currently has over 1 million users. The numbers are increasing, as the PAEHR accounts for 10000-13000 users per day [14].

Methods

Based on the hypothesis that the new version better supports patient-centered care, this study qualitatively analyzed the first (v.1.0) and the second version (v.2.0) of the NRF, and the functionality of the PAEHR. Data collection regarding the NRF was performed via Inera.se website, the owner of Swedish eHealth services [14], whereas data collection of the PAEHR version 2.6.2 was through 1177.se, the national portal

of all public eHealth services in Sweden. The analysis started off with retrospective reviews (managed by IS) based on experiences and knowledge of the other authors (MP, BE, LL). Subsequently, key concepts of the NRFs were thematically analyzed [15] with regard to the following patient-centered care: children; relatives; information handling; harm; and secrecy. Focus was on eliciting the rationale of the changes in the NRF.

All the authors (BE, LL, MP) have extensive experience with the PAEHR and the NRF. They developed the service in Uppsala County Council (UCC), the first region to deploy the eHealth service regionally, as well as its preceding pilots since 1997 [16]. They also designed the framework of rules valid for UCC, and acted as consultants for the creation of the first national framework [17]. The analysis of the current PAEHR and the NRF 2.0 (in operation since October 2016) was, therefore, based on user data as well as experiences regarding usage during the deployment of the service and NRF v.1.0 [7, 13, 14, 16].

This analysis is limited to the NRF; thus, the analysis of which type of information is displayed in the respective county is not presented here, although sets of information shown still differ between various counties which affects the usage. Decisions from the 18 connected regions were compiled by Inera AB and can be found on their web site [14], whether displaying or not displaying the following sets of information: medical notes; diagnoses; forms; log report; prescriptions; maternity care; lab results; referrals; blocked parts of the record; growth curves; cave and medical alerts; vaccinations; care contacts; and psychiatry notes.

Results

National Regulatory Framework Version 1.0

The previous NRF (v1.0) was established by the National Board of Health and Welfare in 2014 after referral to all county councils and stakeholders in Sweden. It was based on the asset of rules developed by UCC to support its deployment of PAEHR, which at the time had been used by over 115 000 users [13]. Thus, the regulation developed for UCC was adopted by the national development of the service, with the aim to create the basis for the NRF. For example, it stated with ‘Healthcare providers allowing direct access to an individual’s patient data will also be responsible for the existence of an assessment system of the parts that require special protection in relation to the individual and shall not be disclosed by direct access’ [13]. That statement, among others, opened up discussions within and between the other counties. Such information entities are per se delicate, and by tradition and culture handled differently by various care providers. Due to the self-governance of the county councils, the NRF 1.0 resulted in a set of both mandatory paragraphs and paragraphs where some content was optional, and where each health authority made decisions for their regional development of the service.

Mandatory paragraphs

1. The individual must be identified by secure login.
2. The individual should be informed about where to turn for help to understand what is written in the EHR, as well as to which extent information in the EHR is *not* presented in the PAEHR.
3. Adults from 18 years should have direct access to their own PAEHR.

4. Guardians have access to their child’s information until the child turns 13 years.
5. Adults shall be able to appoint other adults as being their representatives and give them direct access to the proper PAEHR.
6. Children up to 17 years shall not have direct access to their own health data.
7. Individuals shall be able to seal their PAEHR and shall thereafter not have direct access to their health data.
8. Individuals must be able to read the access log, and thereby, obtain information on healthcare professionals who have opened their EHR and representatives who have opened their PAEHR.

Electable paragraphs

Application of the **electable paragraphs** was decided by each health authority and, consequently, the content of the PAEHR currently varies between the county councils.

9. The individual should opt for the EHR to be available for direct access, regardless if signed/authenticated – or to be given direct access only if it is signed /authenticated.
10. The individual should opt for the EHR to be immediately available for direct access – or to be given direct access with the delay of 14 days.
11. Medical notes categorized by keyword “Early hypothesis” should not be accessible to the individual by direct access.
12. Medical notes categorized by keyword “Exposure to violence in close relationships” should not be accessible to the individual by direct access.
13. Health data not available for direct access: from care units where the health authority, through a policy decision, has determined that manual audit should precede extradition.
14. Health data not available for direct access: from care professions who the health authority, through a policy decision, has determined that manual audit should precede extradition.
15. The individual must receive direct access to all information in the EHR that is electronically available (with exceptions given above) [14]

Proposed rules 11 to 15 could be either accepted or rejected by the various county councils.

National Regulatory Framework Version 2.0

Although connected regions accepted the use of the NRF v 1.0 and submitted their decisions regarding the electable paragraphs, there was a debate on how to apply certain functionality with respect to the optional rules of the NRF. As stated in the analysis of NRF v. 1.0 [13], there was a need to revise the NRF to provide less electable paragraphs, and to require an adherence to the next NRF by all county councils.

On 22 September 2016, the Board of Inera took the decision on a new framework for the PAEHR that all counties and regions should be able to endorse. The new framework has been developed together with the Swedish Association of Local Authorities and Regions (SALAR) and replaces the previous regulations [14]. Based on the goals of the National Action Plan of eHealth [18] and subsequent updates, “all residents from 16 years should by 2020 have access to all information

documented in county-funded health and dental care through the PAEHR 'Journalen'. Using PAEHR, every individual can reach all information about themselves and actively participate in the healthcare. Care providers take responsibility for the information related to their organizational business, but the individual should have full access to it; anywhere, anyhow and anytime" [14].

The next step is to establish this NRF at the political level of management in counties and regions to underpin the continued development and implementation of the PAEHR.

Principles regarding information to be made available:

1a. All digital health records, in the county-funded health and dental care, which may be disclosed to the individual, shall be made available through direct access.

2a. The information shall be made available as soon as it is inserted in the EHR.

3a. The individual shall be able to choose what information he/she wants to see, and whether a selected individual should be invited to read information and to what extent.

4a. Guardians have access to their child's information until the child turns 13 years.

5a. Depending on the situation of the child, in individual cases, it shall be possible to extend or shorten the access of a guardian, as well as to advance the child's own access to his/her information.

The five Available-principles (1a-5a) apply throughout county-funded health and dental care. Individuals are referred to as people from 16 years and all digital medical records are referred to as information supported by Inera's service platform, also historically.

Principles regarding information not to be made available

In exceptional cases and with regard to the law of Public Access and Secrecy, information can be hidden from direct access of an individual. The four principles of exception (1e-4e) concern:

1e. Details of a respondent (third person) in a record entry.

2e. Details of the patient, if it is of particular importance with regard to the purpose of ongoing care and treatment that the information shall not be disclosed to the patient.*

3e. If information may harm a person in exposure to violence.

4e. If information is subject to investigation secrecy at the request of the Police or prosecutors.

*This provision shall apply only in exceptional cases and applies only to patients undergoing care and treatment. A patient with a completed treatment cannot be denied access to the PAEHR.

Analysis and Discussion: Rationale for Change and Implications for the Patient

Compared to the NRF v. 1.0, the second version is based on the national and European action plan of eHealth [18, 19] and, thus, the aim is to deliver more general principles; whereas the first version aimed to support practical implementation of the PAEHR in the Swedish regions based on experiences of the first county to deploy the service. Therefore, the paragraphs were quite hands-on and on a detailed level.


Goal-oriented principles

The major change in the NRF is that it clearly marks that all digital health information shall be made available for direct access of the patient.

Moreover, the first version of the NRF resulted in 21 different interpretations, as each region is autonomous and had the opportunity to select among the electable paragraphs which rules should apply in their region. In the second version, the goals of the European and national eHealth strategies are used to create a number of principles, thereby giving the citizens the same opportunities regardless of where you live, where and when you seek care. Further, it should be possible for the regions to update their solutions to adhere to version 2.0 by 2020, and to incorporate the new goals. One is to include dental care, and another is to show all information that may be disclosed to the individual, via the Inera service platform, using standardized service contracts. The service contracts support municipality data as well as historical data.

Unified principles replace mandatory and electable rules

The 15 paragraphs of version 1.0 are unified into five principles regarding information to be made available, and four exceptions when information should not be exposed to the patient directly (figure 1).



Unifying of the rules to improve alignment of the regions regarding:	15 paragraphs (1-8 mandatory, 9-15 electable)	9 principles (5 info available, 4 exceptions)
Children	3, 4, 6	4a, 5a
Agent reads PAEHR	5	3a
How PAEHR is handled	1 tech, 2 info, 13, 14	1a, 2a, 3a
How EHR is handled	8	
.. All info to patient	.. rule out 9,10,11,15	.. 1a, 2a
Potential harm	7 seal, 12	2e, 3e
.. Harm of patient	.. rule out 13, 14	2e, 3e
Law of secrecy		1e, 4e

Figure 1- 9 Principles of v.2.0 replace 15 paragraphs of v.1.0

In a wide interpretation of the rules it is possible to group the paragraphs of NRF 1.0 and the principles of NRF 2.0. The ones that explicitly handle children and adults (from 16 years) are now described in 4a and 5a. 3a states that individuals should choose what to read and who should be able to read the information. Mandatory paragraphs of v.1.0 regarded technical requirements and information about the service, such as what type of data was not available for direct access, as well as logs of who accessed the EHR or the PAEHR. They are now replaced by the 1a, 2a and 3a principles, which also rule out the electable paragraphs 9, 10, 11 and 15. In the same way, the exceptions 2e and 3e that address potential harm to the patient in different circumstances, rule out electable paragraphs 13 and 14. The mandatory paragraph number 7 addresses the action of sealing the PAEHR, which has no direct equivalent in the new NRF. Alike the old paragraphs 1 (secure log in) and 2 (information and help), paragraph 7 is now handled as a functional requirement. Electable paragraph 12 is replaced with exception 3e. A more clear connection to the law of secrecy is made by adding the exceptions 1e and 4e.

Rationale for Change

The new principles aim to support an alignment and a joint adherence of all county councils and regions in Sweden. Version 2.0 states, in short, that health information should be made available directly regardless of what the healthcare authority previously had decided, as long as the information with regard to the law of secrecy does not risk harming the patient or other persons. Individual rules may apply to children and their guardians.

Principle 1a: The rules for using the service should take into account the patient's demands of the greatest possible access to data. The NRF should not contain options that lead to different accessibility in different parts of the country.

Provision regarding the patients' direct access can be found in Chapter 5, §5 of the Patient Data Act (PDA) [20]: a healthcare provider may allow an individual direct access to data on the individual himself, as may be disclosed to him or her, and treated for purposes specified in Chapter 2, §4 in the PDA. The government justified the provision on direct access as follows: "Giving patients direct access to their medical health data contributes to their ability to a better way to actively participate in their care" (Prop. 2007/08: 126 p. 158). Further, disclosure on different terms gives rise to unequal healthcare and should be avoided.

Principle 2a: In general, the opinion of patients and their representatives has been to show all information immediately, while medical professions have argued that patients should be allowed to have direct access only after a 14 days window [4, 7], as the example of the electable paragraph 10. Principle 2a also refers to electable rule 9: whether the patient should be able to read only signed notes, or get access also to unsigned notes. This issue was settled by the Supreme Administrative Court (HDF 2013 ref. 33). For a public caregiver, record entries are always public documents. This means that such notes are considered established and public before "the ink dries". It is irrelevant according to the court if they are signed or not.

Principle 3a: A person acknowledged by the patient, e.g. a relative or an agent, has the right to take part of the patient's health records through direct access. The Agent functionality in the PAEHR has been tested by the Administrative Court (judgment 2016-06-10, Case No. 5402-15). The court considers the operation legal in accordance with the individual consent under Chapter 2, 3§ PDA; the conditions are the same for the agents as for the individual. The judgment has been appealed.

Principle 4a, 5a: A big change is that the age of majority for accessing health data is now set to 16 years, compared to the previous 18 years. This decision was preceded by discussions with representatives of children, pediatricians and SALAR based on an increasing pressure to lower the majority age for health data, with respect to children's rights not to be excluded. Previously, it was not possible for a 16 years old person to get an e-Identification, as the age of majority in general is 18, and that age seemed reasonable to keep. Now, Sweden follows the example of Norway [6] of trying out the new age of 16 years for accessing health data (prop. 2007/08:126 s. 153). Currently, there is no law that supports this decision as the Parental Code refers to a "degree of maturity", which does not apply for a technical system. However, both the Data Inspection Board and the European Data Protection Regulation (art 8.1) acknowledge that at the age of 16, children can prevail over questions of confidentiality and consent to the processing of personal data.

The guardians lose the possibility to read their child's health information when the child turns 13 years, unless there are individual needs (see 4a, 5a).

Exceptions 1e-4e: The starting point, under current law, is that the patient has the right to take part in all care documentation. In exceptional cases, information can be hidden from direct access of an individual. Compared to NRF v. 1.0, this does not regard information from specific organizations, care facilities or professional groups. The exception addresses confidentiality in relation to the patient pursuant to Chapter 25, §§ 6 and 7 of the law of Public Access and Secrecy.

The NRF v 2.0 in relation to patients' own stories

Introducing PAEHR in Sweden has been a success from the patients' perspective. The PAEHR allows for users to send feedback via email, and the mailbox receives 10 emails a day. Also, the patient portal 1177.se receives feedback from patients and, approximately, 400 phone calls a day, where some are in regards the PAEHR. A brief overview of the feedback content revealed that patients most often had difficulties in understanding the different interpretations of the NRF and the different filtering of the care providers, as a result of that some information is displayed from certain regions while the same information is not displayed from other regions. The most frequent question was the fact that patients want direct access to all information. One user quote illustrated that the PAEHR is not only appreciated by the patients, but also a benefit for the patient in terms of improved health information outcome:

"I have used the eService "Journalen" a few times now, and thanks to the ability to continuously follow up clinic visits, with notes nobody can dispute, after two years, I finally received the referral I was promised already in 2014 and I have taken the tests for the disease that doctors speculated on. I think your service can save healthcare in this country. You deserve a medal and hero status!".

Here the patient benefits from principles 1a, 2a and 3a from NRF v. 2.0, as well as the provision stating that "all digital medical records are referred to as information supported by Inera's service platform, also historically".

Another written remark from a patient illustrated the need of the 4a and 5a principles:

"Hi Journalen! => I am writing from two completely different angles, firstly as myself as a patient. When will more regions be connected to the different parts of the PAEHR? It is amazingly annoying to wait for important test results, to log in, and to see that my region (Skåne) does not report test results. Furthermore, I am a geek (and many with me), and would have had a benefit of arranging the values in, for example in a graph! It is a pity that not everyone uses the great technology that is available. AND: I am the mother of a son who will turn 14 and have a lot of chronic diseases I need to relate to. In the past, I could, before a doctor's appointment log in and read, check the values from last month, remind myself of which preparation he received in his last vaccination etc. From the day he turned 13 there is not this possibility anymore. WHY? Above all: Why is there no agent system for one's own children? My dad can share his medical records – but the one of my son, I cannot have access to? How did you think? I was answered by your support that he can order his records on paper – but that is not at all what we want. This is the WORST I have ever been through, and besides, I think that 13 is a very low age limit. Now it sounded like I am not satisfied with your product, but I am, as a whole. Thanks for letting me leave my comments."

One of the most common complaints of patients is that the latest information cannot be read when the provider has chosen to show only information that is signed or authenticated. Experience shows that positive effects related to direct access outweigh the negative effects [7].

Data analysis of the UCC implementation presented that a majority of patients (98%) chose immediate access to health record information. A window of two weeks is experienced as too long for patients who want to be involved in their care [7]. This is in line with the Swedish eHealth Strategy [18] and the European eHealth Action Plan [19] promoting personalized service and interactive eServices to exercise participation and self-determination on their own terms.

If this transformation of healthcare is going to take place, there is a need to consider development and deployment from a real patient-centered perspective. Although the NRF and the PAEHR now follow the strategy objectives, there is still much work to do to reach a service addressing patient needs to follow his/her care and treatment.

Conclusion

To date, care providers have assessed differently how to apply the NRF v.1.0. The framework of v.2.0 shows potential as it indicates a clear direction towards all information to be made available to the patient. It is more equitable for all patients to get access to and share all their health information regardless of where they live or receive care. The exemption is valid only in specific cases, and will apply only during ongoing care, which means that limits are neither static nor valid forever.

All county councils have endorsed the new framework, and regional decisions to adopt the framework are expected in the near future. The county councils now have a few years to work to get all pieces in place by 2020. However, the increasing usage reveals that citizens probably do not want to wait until 2020 to get involved, and the authors anticipate that the work needs to be intensified, once the ball is set rolling.

Acknowledgements

We would like to thank Inera AB (www.inera.se) for providing the data. The “DOME consortium” studies Deployment of Online Medical Records and eHealth Services” <https://domeprojekt.wordpress.com/> and hosts the project PACESS financed by FORTE (2016-00623).

References

- [1] S.R. Jilka, R. Callahan, N. Sevdalis, E.K. Mayer, A. Darzi. Nothing About Me Without Me: An Interpretative Review of Patient Accessible Electronic Health Records. *J Med Internet Res*;17:6 (2015) e161.
- [2] S. de Lusignan, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. *BMJ Open*;4:9 (2014), e006021.
- [3] G. Erlingsdottir, and C. Lindholm. When patient empowerment encounters professional autonomy: The conflict and negotiation process of inscribing an eHealth service. *Scandinavian Journal of Public Administration* 19, 29 (2015), 27–4
- [4] I. Scandurra, A. Jansson, M-L. Forsberg-Fransson, T. Ålander. Patients' Right to Access their Electronic Health Record Online is Controversial – Lack of knowledge and diverse perceptions among care professions. *International Journal of Reliable and Quality E-Healthcare* 6:1 (2017) 29-45 DOI: 10.4018/IJRQEH.2017010103.
- [5] K.D. Mandl, and I.S. Kohane. Time for a Patient-Driven Health Information Economy? *N Engl J Med* 374 (2016) 205-208.
- [6] T. Sorensen, and M.A. Johansen. Developing and Implementing Patients' Full-Scale Electronic Access to their Health Record, *Stud Health Technol Inform* 228 (2016) 85-89. doi:10.3233/978-1-61499-678-1-85
- [7] L. Lyttkens. Support USERS To Access INFORMATION and Services. Document D6.3 Final Pilot Evaluation Uppsala County Council, v 1.0. EU project report: ICT PSP – Empowering patients, 2015.
- [8] S.G. Shah, R. Fitton A. Hannan, B. Fisher, T. Young, J. Barnett. Accessing personal medical records online: a means to what ends? *Int J Med Inform* 2015, 84(2):111-118.
- [9] T. Delbanco, J. Walker, SK. Bell, JD Darer, JG Elmore, N Farag, HJ. Feldman, R. Mejilla, L. Ngo, JD Ralston, SE. Ross, N. Trivedi, E. Vodicka, SG. Leveille. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. *Ann Intern Med* 157:7 (2012), 461-470
- [10] D. Wiljer, S. Urowitz, E. Apatu, et al. Patient accessible EHR: exploring recommendations for successful implementation strategies. *J Med Internet Res* 10:4 (2008) e34. doi: 10.2196/jmir.1061
- [11] M. Tiik. Rules and access rights of the Estonian integrated e-Health system. *Stud Health Technol Inform*;156 (2010), 245–256.
- [12] L. Beard, R. Schein, D. Morra, K. Wilson, J. Keelan. The challenges in making electronic health records accessible to patients. *J Am Med Inform Assoc*;19:1 (2012) 116-120.
- [13] I. Scandurra, L. Lyttkens, B. Eklund, Implications of Swedish National Regulatory Framework of the Patient Accessible Electronic Health Record. *Stud Health Technol Inform* 228 (2016), 695-699. ISSN 1879-8365. doi:10.3233/978-1-61499-678-1-69
- [14] Inera. *Journalen*. <http://www.inera.se/TJANSTER--PROJEKT/Journalen/> Retrieved 2017-04-13
- [15] B.J. Oates. *Researching Information Systems and Computing*. (2006) London: SAGE
- [16] B. Eklund and I. Joustra-Enquist. Sustains - direct access for the patient to the medical record over the Internet. *Stud Health Technol Inform* 100, (2004) 182-189.
- [17] I. Scandurra, RM Åhlfeldt, Å Cajander. Towards National Deployment of Medical Records and eHealth Services. *VITALIS – Nordic leading eHealth meeting* (2014), 14-17. https://gupea.ub.gu.se/bitstream/2077/35435/2/gupea_2077_35435_2.pdf.
- [18] Ministry of Health & Social Affairs. *National eHealth – the Strategy for Accessible and Secure Information in Health and Social Care*. Retrieved 2016-12-05 www.regeringen.se/sb/d/108/a/148429
- [19] European Commission, eHealth Action Plan 2012-2020 – Innovative healthcare for the 21st century, http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=9156 Retrieved 2016-12-05
- [20] Patient Data Act. *SFS* 2008:355. Swedish Government, http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Patientdatalag-2008355_sfs-2008-355/ Retrieved 2016-12-05

Address for correspondence

Isabella Scandurra, Örebro University, School of Business, Informatics, SE-701 82 Örebro, Sweden;
e-mail: isabella.scandurra@oru.se

Green Turning Brown – Domain Engineering for Social and Health Services in Finland

Reima Suomi^a, Pirkko Nykänen^{a,b}, Tapio Vepsäläinen^a, Riina Hiltunen^a

^a Department of Management and Entrepreneurship, University of Turku, Finland

^b Faculty of Natural Sciences, University of Tampere, Finland

Abstract

Being able to design information systems to an untouched domain, without the burden of existing information systems, especially legacy systems, is often seen as a dream of most information system professionals. Uncharted domains are anyway scarce, and often such greenfield projects turn into brownfield projects, also to projects where existing structures severely constrain the development of new systems. In this article we discuss the concepts of greenfield and brownfield domain engineering and software development, and reflect their possible messages to the re-engineering of the Finnish health- and social care ecosystem currently under way. In our fieldwork we could identify a lot of need and wish for greenfield domain engineering in the Finnish health and social services delivery. As well we found a lot of brownfield elements inhibiting change. Our proposal for the future is a ecosystem approach, where new and established elements could live together in a self-governed balance.

Keywords:

Knowledge Management; Information Systems; Decision Making

Introduction

Health and social care are under restructuring in many countries [1], including well-known programs like Obamacare in the US [2] and reform of National Health Service in UK [3] and Sweden [4]. In Finland, the renewal of health and social services (the Finnish SOTE-reform) is currently under way, and it will demand a lot of new governance structures, processes and supporting ICT infrastructure to become a reality, the whole health and social care ecosystem will deeply reinvent itself in Finland. This puts heavy pressures on many domains, including, and not to the least, information systems in the industry. Whereas all would like to see this new ecosystem as a greenfield design, also a design without restrictions from the past, in reality we talk of a brownfield design: old structures must be taken into account, and they often inhibit the implementation of the best available solutions. Greenfield and brownfield terminology is eagerly used in land usage planning [5; 6], but the metaphors of greenfield and brownfield have also found their way to the field of organizational context [7; 8] and information systems [9].

In this article, we review based on academic literature what brownfield and greenfield development means in organizational and industry context, and what it specifically means in the design of information systems. We look for

insights that could be used when redesigning information systems in health and social care domain.

Total rework of a large entity can be called domain engineering. Domain engineering is often seen as focusing on software. Domain engineering is relevant to the work required to establish a set of software artifacts that can be reused by the software engineer. The purpose of domain engineering is to identify, model, construct, catalog and disseminate a set of software artifacts that can be applied to existing and future software in a particular application domain [10]. There is anyway evidence that domain engineering is more than just software work: Domain engineering aims to support systematic reuse, focusing on modeling common knowledge in a problem domain [11].

The major alternatives for domain engineering are greenfield and brownfield engineering.

Most of the system and software requirements literature assume development of system from scratch i.e. Green Field Systems [12]. New product development or greenfield process (which does not include constraints for development work like brownfield process) has higher risks. Markets usually have dominant designs, which affect the customer behaviour. When a new product has been developed, there is a risk that the customers do not accept it. Investments to infrastructure of the organization and existing resources have an effect on the selection of whether to develop current products to higher level or to develop completely new products [13].

Lehtonen et al. [13] define that the brownfield (process) stands for the reusing of available assets and it includes notions that there are limitations to designing and solutions because of existing structures. Old product solutions, product structures or customer requirements limit designing of new products. Because of this, the brownfield process is not the preferred solution from the designer point of view. Most organizations in this era have existing large-scale or medium-scale operational systems. With the evolvement of new business requirements arise for change in existing systems to meet evolving business demands and needs. In general, the challenges faced by industries are mainly the capture of requirements for changes and re-engineering in operational systems. Unfortunately, there are very limited approaches defined for re-engineering and changes in existing Brownfield operational systems because most of the system & software requirements in the literature assume the development of system from scratch i.e. Green Field Systems [12].

Infrastructure building is a key task in domain engineering: a domain engineering process should encompass at least three main activities: domain analysis, infrastructure specification and infrastructure implementation [14].

A common scenario of brownfield system development can be a major upgrade in current operational system in terms of the following requirements [12]:

- Incorporation of new business rules in existing system
- Adding up new feature in existing system
- Up-gradation of existing feature
- Adoption of new technology
- Legalization/Certification of product or specific feature

Today and increasingly in the future, most large software-intensive system (SIS) developments will be constrained by the need to provide continuity of service while migrating their services away from poorly structured and documented legacy software applications. Yet most SIS process models contain underlying assumptions that an application project starts from scratch in an unconstrained Greenfield approach [15]. We feel that this is a more than core description of health and social care projects in most countries – including Finland.

A health and social care system reform has been on the agenda of several Finnish governments. Prime Minister Sipilä's current government has a plan to put the new health and social care system into operation by 2019. One of the main targets of the ongoing healthcare system reform is to reorganize service providers into larger units called health and social care provider regions. The aim is a full horizontal and vertical integration of health and social care and the primary and secondary levels of services. [16]

New key actors in the new ecosystem will be regions, that will order the health and social care services from the markets, where public, private and third sector service providers compete in equal terms as much as possible. These new regions are yet inexistent, and so are their information systems. They need a totally new ICT infrastructure that is related to and built on the currently available one, but especially systems to manage contracts with service providers, and systems to foresee and plan for service demands, so that they can purchase the right amount and quality of services from the markets. Similar systems are rather unseen in the Finnish context at least; here we are not speaking of ordinary procurement systems [17]. One of the domains where public authorities have most experience of purchasing services might be logistics [18; 19].

Messages from Brown/Greenfield development to health- and social care work

The problem area of Brownfield system development seems to be rather sparsely understood in health and social care information systems field. Much more understanding needs to be gained on how existing structures inhibit the implementation of best new practices. In healthcare, especially the ponderous patient records systems are a core element, to which all new systems must adjust. As well different other systems, such as different coding systems and terminologies inhibit change. On top of everything else, the industry is heavily regulated.

As said, domain engineering is about modelling common knowledge in a domain. Knowledge in the area of medicine especially is extremely extensive, and there is very little room for greenfield engineering or modelling. Yet new ecosystem functions might necessitate a fresh view on knowledge too.

As a reprieve to the problems, IS projects meet with brownfield elements Boehm [15] proposes incremental commitment building. All decisions concerning the new systems are not done at once, and they are not done by just one party. In order to rapidly and successfully adapt to the increasing rates of change, projects need to be able to concurrently, rather than sequentially, assess and manage (1) opportunities and risks; (2) requirements, solutions, plans, and business cases; and (3) hardware, software and human factors [15]. This kind of approach is clearly needed and to a great extent also visible in the Finnish renewal of health and social care.

Brownfield development means that there are established structures, experts and solid knowledge of the domain area already available. Careful listening to experts and taking their ideas into attention and production is a key success factor for brownfield projects. It is important to remember that customers and ordinary staff members are often the best experts in details, even when they are not always able to structure the total picture. The chance of success for Brownfield domain engineering is greatly enhanced if the people working together are effective communicators, like working in teams, are humble, and enjoy learning from others. All of these are important to help create a "no blame" culture, where people aren't afraid to help one another and do what is right for the project [9].

Infrastructure development is a key action in domain engineering. Again, infrastructure in health and social care might in many cases not be the most sophisticated one, but it is usually deeply rooted and difficult to change. Different standards of healthcare message exchange are at the core of infrastructure development in health and social services [20].

Table 1 summarizes our main findings from the literature on brown/greenfield engineering.

Table 1- Main findings from the literature on brown/greenfield engineering

- | |
|---|
| <ul style="list-style-type: none"> • Greenfield domain engineering is always more risky than brownfield domain engineering • Infrastructure planning (architecture) and implementation are at the core of domain engineering • Continuity of service is a major goal • Re-use of current resources and assets is a goal of brownfield domain engineering • Listen to experts of the old system • In IS, greenfield engineering aims migrating services away from poorly structured and documented legacy software applications • Incremental commitment building is important in brownfield domain engineering |
|---|

Methods

Our fieldwork took four months in December 2016 - March 2017 in Southwest Finland. During that time we held 12 expert interviews with Finnish experts in social and healthcare information systems, and one workshop. The theoretical background of green or brownfield domain engineering was not present in the empirical discussions. Our study approach was explorative [21].

Results

In our fieldwork we could identify a lot of pressures and change in agents to the current situation. Genuine interest to design the whole information governance in Finnish health- and social care clearly exists. The main identified factors are summarized in Table 1.

The most clear brownfield element in Finnish health and social care domain of health and social care information governance is the established portfolio of many clinical healthcare information systems. While the amount of different system brands can be counted in thousands, a few main systems – especially patient record systems – account for the main part of current system portfolio. The new and central Finnish main database for health and social care customer information Kanta is on the edge – it can be interpreted as a new building block in an old brownfield landscape.

The current situation in health and social care IS can be described in Finland as waiting. A lot of planning work goes on, but real investments into new systems are few, because of the unclear situation of the upcoming new health and social care governance. It is still widely unclear, what kind of organizations will operate in the new domain, and with which kind of processes and business models, and which kind of incentives are given to different players by government and market forces.

In Finland, there is already a lot of hefty discussion on the future of the fixed property, which will be most likely deserted in the new market environment – especially facilities in rural areas are in danger. Less discussion has been on the same possible fate of heavy investments in current information systems and infrastructure.

Finally, making Greenfield domain engineering is made difficult because of very traditional reasons: lack of finance, lack of vision and direction, and lack of market or other new alternatives that would really look like superior ones as compared to current status.

Table 2 - Main freezers and change agents for information governance in Finnish health- and social care renewal

Main freezers – brownfield elements
<ul style="list-style-type: none"> • Established patient record systems with high market share • Unclear governance • Unclear processes • Unclear incentive systems and business models • Heavy investments in current system portfolio and infrastructure • Lack of finance • Lack of vision and direction • Lack of superior market offerings
Main change agents – greenfield elements
<ul style="list-style-type: none"> • Ambitious political goals • Ambitious architecture plans • New regions as actors • Need for decision support • Need for secondary data use • New patient record initiatives • Artificial intelligence, IBM and Watson

Political agenda setting in Finnish health and social care field has been ambitious. However the situation that there would be

a lot of fresh resources and areas to untap, which is a key idea in greenfield domain engineering, is missing.

There also seems to be going on some rich and ambitious work on new IS architectures for the health and social care domain. Unfortunately, however, the work seems to form an archipelago; the bridges joining the work items are few and far.

The new regions should act as main change agents. Their capacity and capability to do so is still under speculation, as they are not established yet. The goal is to have strong and wisely led regions as a major change agents for Finnish health and social care domain and its information systems.

There seems to be a strong consensus that especially decision making in the new health and social care domain needs strengthening. Decision makers at all levels should be fed with relevant and fresh information. The same holds true for secondary use of information in research and development domains.

We can also see some new initiatives in the main system portfolio. The nation-wide Kanta-service was already mentioned. In addition, some regions plan for ambitious fresh patient records systems and related functionalities. IBM seems to enter the field strongly with its artificial intelligence Watson concepts, that it seems to utilize as a Trojan Horse to more strongly access the Finnish health and social care IS market.

Discussion

A promising approach for governance and management of the complexity, wide scope and brown/greenfield development of the social and healthcare system is the ecosystem approach. An eHealth ecosystem is composed of healthcare organizations, both public and private, service provider professionals, customers, citizens and patients, industrial companies providing their products and services, and technology-mediated communication and infrastructures that in collaboration provide add-on value for both service consumers and other service providers. Infrastructures and networks are needed both for knowledge sharing and management, and for exchanging and communicating information and data. In an ecosystem, services and products are developed and delivered to fulfill the customers' needs, or regulation-stated needs, and the role of marketing and commercialization is minor [22]. Customers are both health professionals and patients, citizens, and service providers are healthcare organizations and various suppliers and industrial companies to offer systems and services to be used by healthcare organizations and patients. This approach supports very well the regional model of the Finnish social and healthcare reform. Good examples of ecosystem applications and services are electronic health records, personal health records, patient portals and health information systems including health knowledge management and e-learning for healthcare professionals. Further, these systems include clinical decision support systems and remote patient monitoring and management applications to be used at home and on-move by patients and citizens. Wellnes and fitness applications should also be mentioned.

An eHealth ecosystem is very dynamic system that incorporates a varying number of stakeholders. Ecosystem requires enabling information, communication, and empowerment mechanisms, which make it possible for information and expertise to be accessed quickly and

accurately to inform and guide the ecosystem activities and performance. An eHealth ecosystem is very specific in the sense that political decision making has strong effects on how the healthcare services are organized and funded, and also, there is strong legal and normative regulation on how the services can be delivered, accessed, disclosed, charged and funded [23]. Typical for an eHealth ecosystem today is that the services, e.g. electronic health records, need to be delivered online, across distinct organizational, regional or even national borders. This requires that an ecosystem is composed of interconnected stakeholders, each one with a mission to improve the quality of care. In this situation, in order to ensure the patient safety and quality care, the stakeholders build new relationships, often outside the healthcare organization, e.g. develop new types of public-private partnerships [24].

The ecosystem approach on eHealth systems and applications, especially in this new nation-wide social and healthcare reform context, helps healthcare organizations and e-health stakeholders to create business models based on collaborative service production and thus improve collaboration and communication between the stakeholders. The actors, partners of the ecosystem need to share the common objectives of the system and benefit from collaboration, create shared value, and when they have adequate tools and means for communication, collaboration will be active and support the ecosystem sustainability.

eHealth ecosystem can provide many potential benefits for the healthcare professionals and organizations and patients and citizens and for industrial suppliers. However, creating sustainable eHealth ecosystems requires that all stakeholders' opinions and needs are taken into account for ecosystem success and sustainability. Sustainability is important because the benefits often can be achieved only over a period of time. This kind of approach is very relevant, essential, and necessary for the successful governance and management of our planned social and healthcare reform.

Conclusions

Our explorative journey confirmed that terminology and thinking on greenfield and brownfield domain engineering fits well even to the social and healthcare domain information systems development. This to a great amount in health and social care, IS development untapped theoretical construction, and stream of literature and research offers new avenues and insights to research in health and social care IS.

In reality healthcare is plagued with brownfield elements that hamper new openings of IS in the domain. Social care is having less past burden, especially in IS governance, but even there old traditions might heavily hamper new development ideas.

A lot of long-term IS planning in the form of architectures is taking place in the Finnish health- and social care domain. We can easily see that the development initiatives are forming an archipelago, with few bridges between the different plans and a lot of overlapping work.

Incremental commitment building is a key to success in brownfield domain engineering projects. Listening to different stakeholders and allowing them to make their real input to the future IS governance arrangements is an enabler of success, but of course not yet a silver bullet. The cost of this approach is most likely a lot of superimposed work, but this is the cost that has to be paid for consensus. Allowing wide groups of

stakeholders to co-produce the plans also serves educational and learning purposes.

Our proposal for a fresh approach is more underlining the nature of an ecosystems of the health and social care domain. Rational top-down planning will never harness the realities of the domain – even not in a small country like Finland. Rather we need more focus on the understanding on how the health and social care ecosystem works, and how IS as a sub-ecosystem in the area works.

References

- [1] C. Ham, Health care reform: learning from international experience, *Nursing Management* 4 (1997), 26-30.
- [2] M.A. Hall and R. Lord, Obamacare: what the Affordable Care Act means for patients and physicians, *British Medical Journal* 349 (2014), 5376-5376.
- [3] S.P. Bate and G. Robert, Knowledge management and communities of practice in the private sector: lessons for modernizing the National Health Service in England and Wales, *Public Administration* 80 (2002), 643-663.
- [4] H. Hasselblad and E. Bejerot, Performative policy: the case of Swedish healthcare reforms, *Critical Policy Studies* (2016), 1-20.
- [5] D. Lange and S. McNeil, Clean it and they will come? Defining successful brownfield development, *Journal of urban planning and development* 130 (2004), 101-108.
- [6] R. Ganser and K. Williams, Brownfield development: Are we using the right targets? Evidence from England and Germany, *European Planning Studies* 15 (2007), 603-622.
- [7] E.H. Schein, How Can Organizations Learn Faster? The Challenge of Entering the Green Room, *Sloan Management Review* 34 (1993), 85-92.
- [8] F.E. Emery, Designing socio-technical systems for 'greenfield' sites, *Journal of Occupational Behaviour* 1 (1980), 19-27.
- [9] J. Rasmussen, Introducing XP into greenfield projects: lessons learned, *Ieee Software* 20 (2003), 21.
- [10] R.S. Pressman, *Software engineering: a practitioner's approach*, McGraw-Hill, New York, 2000.
- [11] R.d.A. Falbo, G. Guizzardi, and K.C. Duarte, An ontological approach to domain engineering, in: *Proceedings of the 14th international conference on software engineering and knowledge engineering*, ACM, 2002, pp. 351-358.
- [12] A. Masood and M.A. Ali, Applying Agile Requirements Engineering Approach for Re-engineering & Changes in existing Brownfield Adaptive Systems, *arXiv preprint arXiv:1410.6902* (2014).
- [13] T. Lehtonen, J. Pakkanen, M. Lanz, J. Järvenpää, and R. Tuokko, Brownfield Process for Developing of Product Families, in: *DS 68-4: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 4: Product and Systems Design*, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011, 2011.
- [14] G. Arango, Domain Analysis - From Art Form to Engineering Discipline, in: *The 5th International Workshop on Software Specification and Design*, ACM, Pittsburg, 1989, pp. 152-159.
- [15] B. Boehm, Applying the Incremental Commitment Model to Brownfield Systems Development, in: *7th Annual Conference on Systems Engineering Research*, Loughborough, 2009.
- [16] H. Hyppönen, P. Hämäläinen, and J. Reponen, eds., *E-health and e-welfare of Finland - Check point 2015*, National Institute of Health and Welfare, Helsinki, 2015.
- [17] C. Chang, An Empirical Investigation of Procurement System Selection: A Transaction Cost Explanation, in: *IEEE International Conference on Engineering Management and Service Sciences (EMS 2010)*, Beijing, China, 2010.
- [18] G. Krawczyk, The Impact of Public Procurement System on Concentration of the Rail Freight Market in Poland, in: *Transport Development Challenges in the Twenty-First Century*, Springer, 2016, pp. 31-38.
- [19] L. Borowik and A. Cywiński, Modernization of a trolleybus line system in Tychy as an example of eco-efficient initiative towards a sustainable transport system, *Journal of Cleaner Production* 117 (2016), 188-198.
- [20] C.N. Mead, Data interchange standards in healthcare it-computable semantic interoperability: Now possible but still difficult. Do we really need a better mousetrap?, *Journal of Healthcare Information Management* 20 (2006), 71.
- [21] J.W. Creswell and V.L.P. Clark, *Designing and conducting mixed methods research*, SAGE Publishing., 2007.
- [22] V. Vimarlund and T. Mettler, Introduction to the ecosystem for two-sided markets, barriers and facilitators, in: *E-health two-sided markets – Implementation and business models*, V. Vimarlund, ed., Elsevier Academic Press, 2017, pp. 3-18.

- [23] P. Nykänen, Implementation and evaluation of ehealth ecosystems in two-sided markets, in: *E-health two-sided markets – Implementation and business models*, V. Vimarlund, ed., Elsevier Academic Press, 2017, pp. 125-142.
- [24] T. Bibin, The new healthcare ecosystems – 5 emerging relationships, in: *Becker's Hospital Review*, ASC Communications, 2014.

Address for correspondence

Professor Reima Suomi
University of Turku, Finland
Email: reima.suomi@utu.fi

Toward Automatic Reporting of Infectious Diseases

Hsiao-Mei Tsao^a, Chi-Ming Chang^b, Jen-Hsiang Chuang^{b,c},
Ding-Ping Liu^b, Mei-Lien Pan^a, Da-Wei Wang^a

^a Institute of Information Science, Academia Sinica, Taipei, Taiwan, China

^b Epidemic Intelligence Center, Centers for Disease Control, Taipei, Taiwan, China

^c Institute of Biomedical Informatics & Institute of Public Health,
National Yang-Ming University, Taipei, Taiwan, China

Abstract

Accurate, complete, and timely disease surveillance data are vital for disease control. We report a national scale effort to automatically extract information from electronic medical records as well as electronic laboratory systems. The extracted information is then transferred to the centers of disease control after a proper confirmation process. The coverage rates of the automated reporting systems are over 50%. Not only is the workload of surveillance greatly reduced, but also reporting is completed in near real-time. From our experiences, a system sustainable strategy, well-defined working plan, and multifaceted team coordination work effectively. Knowledge management reduces the cost to maintain the system. Training courses with hands-on practice and reference documents are useful for LOINC adoption.

Keywords:

Public Health Surveillance; Electronic Health Records;
Logical Observation Identifiers Names and Codes

Introduction

Infectious disease surveillance has been a cornerstone of disease control. In order to have effective surveillance systems, we need to obtain timely, accurate, and complete data. In the past, data collection process used to rely on the physicians and Infection Control Nurses (ICNs) to manually report via telephone or faxing to the health authorities. The manual reporting process is not only time-consuming and labor-intensive, but also error-prone. With the adoption of web-based information systems, reporting time has greatly reduced and data quality has significantly improved. Information systems have transformed from a supportive role for public health to a proactive one in the last two decades. In 1995, the first syndromic surveillance systems were established to detect outbreaks of waterborne illness [1]. In addition, the US government started promoting a National Electronic Disease Surveillance System (NEDSS) in 2000. Hospitals were encouraged to develop electronic reporting systems to achieve stage 1 of Meaningful Use [2]. In 2004, the Public Health Information Network (PHIN) integrated the information systems to meet public health functional needs [3]. These all improved case detection and lead to prompt responses to an outbreak.

Currently, the major surveillance systems in Taiwan are the notifiable disease reporting system, laboratory surveillance system, school surveillance system, populous institutions surveillance system, and Taiwan Real-time Outbreak and Disease Surveillance (Taiwan RODS) system [4]. Among them, the notifiable disease system is the most important one.

Physicians and ICNs are mandated to report notifiable diseases through this web-based system. That is to say, they need to look up all the information from hospital information systems and then type information into the notifiable reporting system. According to Taiwan's regulations, there are more than seventy notifiable diseases. For each disease, plenty of data needs to be collected. Thus, collecting surveillance data comprises heavy workloads for physicians and ICNs. In addition to the notifiable disease system, Taiwan RODS is also important for infectious diseases surveillance. Taiwan was severely struck by the SARS epidemic, and subsequently, the government has made a lot of progress in disease surveillance. Taiwan RODS was launched to automatically monitor several syndromes through emergency departments since 2004. Taiwan RODS was easily adapted to monitor an outbreak of red eye syndrome and has shown its effectiveness in situational awareness [4]. Now, it monitors influenza-like illnesses, enterovirus infections, red eye syndrome, and diarrhea. However, before conducting this project, Taiwan had not yet implemented any electronic laboratory reporting (ELR) system. We could only obtain pathogen information from sentinel physicians through the laboratory surveillance system to characterize the subtypes of influenza virus and enterovirus infections.

In the past decade, the Taiwanese government has aggressively promoted Electronic Medical Record (EMR) systems and Health Information Exchanges (HIE). There are 406 out of 501 hospitals that have implemented EMR systems and are capable of interchanging EMRs among institutions [5]. Given this basis, Taiwan Centers for Disease Control (TCDC) launched the National Epidemic Prevention and Control (NEPC) project to improve the efficiency and effectiveness of surveillance in 2014 [6]. This project comprises two programs. One is the "automated reporting via EMR" program (EMR program for short), which aims to accomplish the automatic machine-to-machine reporting of notifiable diseases from hospitals' EMR systems to TCDC. The other is the ELR program, the goal of which is to report positive laboratory test results of 20 selected pathogens from hospitals (directly from the laboratory information system). In this article, we report Taiwan's experiences of its implementation strategies and current achievements in promoting the NEPC project.

Methods

Implementation strategies

The initial step of the NEPC project was amending the regulation by the authority to make automatic extraction and reporting acceptable. A committee was formed to design the system architecture and establish program workbooks that

contain all the necessary information to implement such systems. The data format as well as the semantics of required data elements of each notifiable disease and an accompanying specification using HL7 [7] Green Clinical Document Architecture (GCDA) [8] were also included.

The three-year NEPC project took an expanding-and-maintaining strategy in three rounds. For each round, which took one year, there were three stages: the recruitment stage, development stage, and evaluation stage. The first two months were the recruitment stage. TCDC announced the latest program workbooks and recruited hospitals. The incentives for hospitals to join the project included funding for system construction and bonus points for hospital accreditation. The recruitment stage ended with a signed contract, and the seven-month development stage began. During this stage, the participating hospitals followed the program workbooks to develop the reporting system and had to pass the system testing at the end. The last three months were evaluation stage. Data completeness and system stability were evaluated at this stage. In order to encourage the participating hospitals to report complete data and maintain system stability, an extra bonus was given based on the results of the evaluation stage. It is worth mentioning that hospitals joining in earlier rounds were actively engaged and this was the maintenance art of the strategy.

In order to run the project smoothly, TCDC organized a multifaceted team, which consisted of a policy instruction committee, an information technology (IT) support group, and a project coordination office. The team meeting was held weekly to monitor the progress and to assist all the hospitals for implementation issues. The details of the development and evaluation stages for the EMR and ELR programs are as follows.

EMR program

The EMR reporting system is designed to utilize hospital EMR to report notifiable diseases. A schematic view of the EMR reporting system is shown in Figure 1. We selected participating hospitals according to the reporting quantity. The participating hospitals have to develop a reporting module, XML generation module, and digital signature module. The reporting module extracts cases from EMR systems; the XML generation module creates reports conform to the defined XML schema; the electronic signature module adds digital signatures to the reports. After physicians or ICNs review the data, the hospitals report the cases to TCDC. For each hospital, we compute a weighted sum of the ratios of all notifiable diseases cases reported by that hospital divided by the national total and this is the coverage rate of the hospital. We can then add all the coverage rates of all participating hospitals as the coverage rate of the project. Furthermore, system maintenance is estimated by the EMR reporting rates and successful reporting rates. The reporting rates of each participating hospital have been monitored since the systems were launched. We also surveyed how frequently the participating hospitals send the reports.

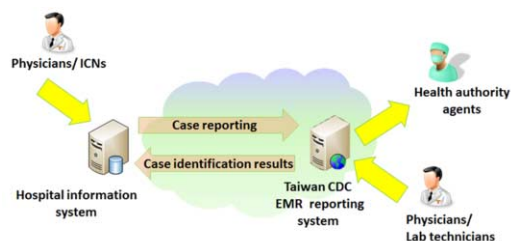


Figure 1—Schematic view of the EMR reporting system

ELR program

The Laboratory Automated Reporting System (LARS) is designed to monitor trends in 20 selected pathogens such as hepatitis virus, influenza virus, and Salmonella, to name a few. A schematic view of the LARS is shown in Figure 2. There are two objectives when selecting participating hospitals: the reporting quantity, and balance of area. The participating hospitals were asked to send positive laboratory test results to TCDC. There are three steps for system development, including gateway setup, Logical Observation Identifiers Names and Codes (LOINC) [9] mapping, and data transmission. The IT support group assisted the hospitals to setup the gateway. The positive laboratory test results were then sent to TCDC. Because there is no reference data for LARS, we used the notifiable disease reporting rates of all hospitals participating in the ELR program in 2015 as a proxy for the coverage rate. The total coverage rate and the district coverage rates were calculated. (Taiwan is divided into six districts according to geographic location.) We adopted the international standards for the laboratory test results, LOINC, as well as the coding system for LARS. Therefore, the first step is to map the hospital local codes to LOINC codes. When the project was initiated, LOINC was not commonly known in Taiwan. Medical technologists (MTs) were not familiar with LOINC. TCDC provided LOINC mapping training courses to all the participating hospitals on the second month of the development stage in each round. The courses contain the introduction of basic knowledge of LOINC, mapping procedure demonstration, and hands-on practice. RELMA[®] [10] was chosen as the mapping tool.

In addition to the training courses, TCDC keeps auditing data to ensure the data quality and the auditing results are fed back to the hospitals. The six parts of LOINC are compared with the mapped LOINC codes to see whether they are matched to each other. The LOINC mapping rates are then calculated. LOINC mapping rate is also one of the items in auditing.

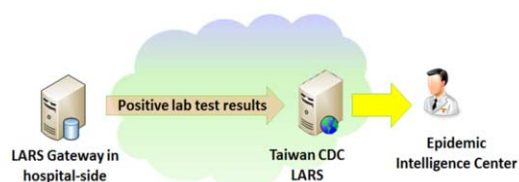


Figure 2—Schematic view of LARS

Results

There are now a total of 47 and 53 participating hospitals in the EMR and ELR programs and the coverage rates are 52% and 59%, respectively. The characteristics of the participating hospitals are shown in Table 1. In the first year, medical centers had higher priority. In the end, all the medical centers in Taiwan joined this project.

Table 1—Hospital levels of the participating hospitals

Program	Hospital level	2014	2015	2016
EMR	Medical center	14	3	2
	Regional hospital	6	6	16
ELR	Medical center	15	2	2
	Regional hospital	5	7	22

EMR program

The implementation strategies of all the EMR program participating hospitals can be categorized into self-developing and outsourcing. Most medical centers (84%) developed their own EMR reporting systems. On the contrary, 75% of the regional hospitals tended to outsource to vendors (Table 2). We conducted a chi-square test with Yates' continuity correction to compare the implementation methods between the two groups and the finding revealed a significant difference with $p < 0.001$. These regional hospitals are served by four health information companies, which is an advantage to the project because the vendors can serve many hospitals.

From our experience, we found that self-developing hospitals were more likely to fall behind schedule. Every year, there are always 1 or 2 hospitals having trouble keeping up with the schedule. Thus, we set up milestones for program progress and carefully monitored the progress of every hospital. Any hospital that was 30 days behind schedule was arranged to be met by the project coordination office. Moreover, it takes a longer time for the self-developed systems to revise the program. This April, the case definition of syphilis was amended; it took about one month to update the program for the self-developed systems, whereas less than two weeks was enough for the outsourced, developed systems to update the program.

Table 2—Characteristics of the EMR program participating hospitals

Hospital level	Self-developed (%)	Outsourced developed (%)
Medical center	16 (84)	3 (16)
Regional hospital	7 (25)	21 (75)
Total	23	24

The EMR program has achieved three marked effects: high reporting rates, near real-time reporting, and decreased workloads.

1. High reporting rates:

All the EMR program participating hospitals were monitored for the EMR reporting rates and successful EMR reporting rates. The average EMR reporting rates were constantly maintained around 93%. Some hospitals had a reporting rate below 60% because they did not develop a reporting module for diseases that rarely occur there. Hospital 41 (H41) is a small hospital. It missed one case, and had only one reporting failure after the system was launched at the end of this September.

2. Near real-time reporting:

The EMR reporting system can automatically extract data from EMR systems. After ICNs examine the reporting data, the data can be sent. Among these participating hospitals, 25 out of 47 hospitals can send data in real-time. The others send data via a scheduled program. The reporting intervals are shown in Table 3. In the future, with enough evidence, the workflow can be further streamlined; the confirmation of the ICNs can be a parallel process with the machine-to-machine reporting. Undoubtedly, the status data will be classified as to-be-confirmed and confirmed. This modification can partially mitigate the potential issue that ICNs become the bottleneck in the process.

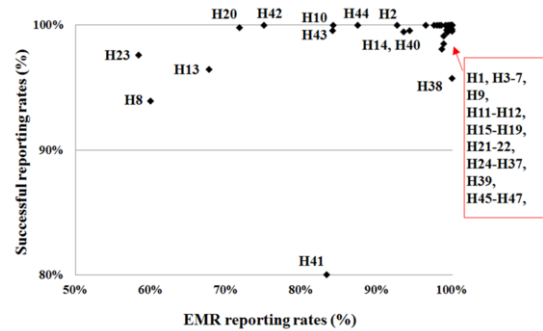


Figure 3—Reporting rates and successful reporting rates of the EMR program participating hospitals

Table 3—Reporting intervals of the EMR program participating hospitals

Reporting interval	No. of hospitals (%)
Real-time	25 (53)
<4 min	15 (32)
4-10 min	6 (13)
11-30 min	1 (2)
Total	47 (100)

3. Decreased workloads:

The most important benefit of the EMR reporting system is the reduction of ICNs' workloads. Because the ICNs' role changes from entering the data to reviewing them for confirmation only, it reduces the reporting time from 8 minutes per case to 3 minutes, as mentioned by a senior ICN from a medical center [11].

ELR program

Compared with the traditional laboratory surveillance system, which collects data from only a few contracted laboratories, LARS gathers data directly from hospital laboratory information systems. On average, TCDC received around 13,000 positive laboratory tests of the 20 selected pathogens per week from the 53 participating hospitals. The coverage rates of LARS for each of the six districts ranged from 44%-69%.

From our experience, with the help of knowledge management, LOINC mapping training courses are the most effective way to promote LOINC adoption. The average LOINC mapping rate was about 50% in 2014, and has improved steadily since. It has now reached 82% in 2016. In the first year, only basic knowledge and RELMA demonstration were provided in the training courses and the average mapping rates were low. More efforts were made to correct the data. Therefore, we added hands-on practice in the second year and the average LOINC mapping rates increased to 72%. In addition, we extracted knowledge from data auditing and reached a consensus for LOINC mapping rules from expert panel discussions. The knowledge was documented as LOINC mapping rules, the LOINC mapping FAQ, and the most frequent LOINC combinations. These documents became the teaching materials in the third year.

Here, we used the influenza virus data as an example to represent the system effects of LARS. Figure 4 shows the number of positive influenza virus tests during the period from the 40th week, 2015 to the 48th week, 2016. In this figure, it

clearly shows a peak spanning from the 2nd week, 2016 to the 16th week, 2016. Type A influenza virus accounted for the most positive test results. LARS can detect an unusual signal two weeks earlier than traditional laboratory surveillance systems (data not shown). Based on the pathogen data collected by LARS, TCDC was able to make a timely decision and respond promptly for flu control.

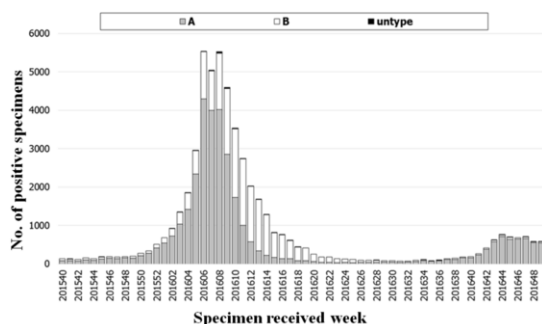


Figure 4—Positive influenza virus tests collected by LARS

Discussion

The NEPC project was successfully executed under a limited funding budget. TCDC has made the coverage rates of both programs reach more than 50%. Besides, the implementation of the EMR reporting systems has greatly reduced the workloads of ICNs. Thus, the reporting rates have been constantly maintained above 90%. The LARS collects positive test results of the selected pathogens and provides further information for infection control efficiently. The experience of the NEPC project provides an operational roadmap for implementing a successful strategy for enhancing national infectious disease surveillance systems, which adopt HIE. From our experience, we have extracted the following three lessons to share.

1. System sustainability:

TCDC makes a tremendous effort to make the operation sustainable. From a policy perspective, the first step is to have a legal basis and then bring in incentives. Hence, the Infectious Disease Control Act and Implementation Regulation for Epidemic Surveillance and Alert System were amended. Thereafter, a one-year grant jumpstarted the program. Further, the maintenance strategy included providing feedback to the hospital and awarding outstanding hospitals. Hospitals also can earn extra bonus points during hospital accreditation. From an implementation perspective, knowledge management is an effective way to maintain the system with limited resources. In this case, we documented each of the issues encountered during the project. As for TCDC, it helps them to provide continuous support to the hospitals. As for the participating hospitals, it is also useful when the personnel change. In addition, the EMR reporting system directly benefits ICNs by reducing their workloads. This makes hospitals more than willing to participate in this project. We have heard that several hospitals were actually persuaded by their infection control departments. Some of the healthcare networks even voluntarily implemented the EMR reporting system in their member hospitals without any funding.

2. Project implementation:

A well-defined working plan and schedule management help the project run on schedule. A full-year detailed working plan and schedule were determined at the beginning with a thorough communication process. The project coordination office monitors the progress in accordance with this plan closely, in order that any delay can be detected earlier. The noted delay triggers a coordinated effort to speed up the progress. Therefore, all of the participating hospitals can finish system construction, testing, and evaluation on schedule.

Furthermore, keeping the project running smoothly requires multifaceted team coordination. There are multiple stakeholders in the project: the hospital, IT support group, project management office, and TCDC. We note that when hospitals need help, the best approach is to have all stakeholders work together to resolve issues. We note that cases in which only IT support group is involved may focus on technical details regardless of administrative support issues, and thus, delay the progress. It seems to be a trivial observation, but when there are several hospitals asking for help, it soon becomes a daunting effort. Furthermore, bilateral efforts soon become the norm if multilateral collaboration is not mandatory. Besides, we note that the needs of the hospitals vary dramatically. For example, medical centers usually have their own in-house software development teams, and the trouble usually comes from inter-departmental communication, especially between the infection control departments/laboratory departments and software teams. The project also has to compete with other projects in the hospital; therefore, an early warning to the project leader, usually the deputy director of the hospital, is very important. For hospitals that do not have in-house software teams, the trouble usually stems from the interaction between two IT vendors, one from the hospital and one from the CDC. The detection of early warning signs and timely intervention from the project office are vital.

3. LOINC adoption:

Although training courses are an effective method to introduce LOINC, only teaching basic knowledge of LOINC is not enough. Hence, we held two expert panel discussions and reached a consensus for LOINC mapping. Thus, overall LOINC mapping rules have been created and incorporated into teaching materials for the next training courses. Online videos for LOINC mapping were also provided. Given the clear instructions and online videos, MTs are able to map LOINC effectively. Because laboratory data change with time, continuous auditing is important. We extracted knowledge from data auditing, and documented the mapping FAQ and the most frequent combinations of LOINC parts.

Conclusion

The NEPC project's success is two-fold: 1) the EMR reporting system greatly reduces the workloads of ICNs and shortens the reporting time of notifiable diseases; 2) LARS collects positive test results of 20 selected pathogens to complement the existing surveillance systems. These achievements have made the surveillance more efficient and effective. From our

experiences, we have learned lessons for system sustainable strategy, project implementation, and LOINC adoption. A system sustainable strategy includes having a legal basis and adequate incentives. Knowledge management reduces the cost to maintain the system. Further, a well-defined working plan, schedule management, and multifaceted team coordination work effectively for project implementation. Training courses with hands-on practice and reference documents are useful for LOINC adoption. Reference documents include the overall mapping rules, mapping FAQ, and the most frequent combinations of LOINC parts.

Acknowledgements

This work was supported by Taiwan CDC (MOHW104-CDC-C-114-000801, MOHW105-CDC-C-114-000401), and by Multidisciplinary Health Cloud Research Program: Technology Development and Application of Big Health Data. Academia Sinica, Taipei, Taiwan. The authors declare that there is no conflict of interests regarding the publication of this article.

References

- [1] R. Heffernan, F. Mostashari, D. Das, M. Besculides, C. Rodriguez, J. Greenko, L. Steiner-Sichel, S. Balter, A. Karpati, P. Thomas, M. Phillips, J. Ackelsberg, E. Lee, J. Leng, J. Hartman, K. Metzger, R. Rosselli, and D. Weiss, New York City syndromic surveillance systems, *MMWR Suppl* 53 (2004), 23-27.
- [2] C. Centers for Disease and Prevention, Progress in improving state and local disease surveillance--United States, 2000-2005, *MMWR Morb Mortal Wkly Rep* 54 (2005), 822-825.
- [3] J.W. Loonsk, S.R. McGarvey, L.A. Conn, and J. Johnson, The Public Health Information Network (PHIN) Preparedness initiative, *J Am Med Inform Assoc* 13 (2006), 1-4.
- [4] J.H. Chuang, A.S. Huang, W.T. Huang, M.T. Liu, J.H. Chou, F.Y. Chang, and W.T. Chiu, Nationwide surveillance of influenza during the pandemic (2009-10) and post-pandemic (2010-11) periods in Taiwan, *PLoS One* 7 (2012), e36120.
- [5] Y.C. Li, J.C. Yen, W.T. Chiu, W.S. Jian, S. Syed-Abdul, and M.H. Hsu, Building a national electronic medical record exchange system - experiences in Taiwan, *Comput Methods Programs Biomed* 121 (2015), 14-20.
- [6] H.-M. Tsao, M.-L. Pan, E.-W. Huang, and D.-W. Wang, Experience from National Medical Informatics Policy: Using the Example of Epidemic Prevention and Control Project, *Journal of Taiwan Association for Medical Informatics* 25 (2016), 1-12.
- [7] HL7 International, Introduction to HL7 Standards, Available from <http://www.hl7.org/implement/standards/index.cfm?ref=nav> (2016).
- [8] R. Worden, Introduction to Green CDA, Retrieved from <http://docslide.us/documents/introduction-to-green-cda-cda-forum-17-th-january-2013.html> (last accessed 16 Dec, 2016).
- [9] Regenstrief Institute, LOINC, Available from <http://loinc.org> (2016).
- [10] Regenstrief Institute, RELMA, Available from <http://loinc.org/downloads> (2016).
- [11] iThome, CDC launched the NEPC project, Available from <http://www.ithome.com.tw/news/94172> (2016) (in Chinese).

Address for Correspondence

Da-Wei Wang

Address: 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan (Institute of Information Science, Academia Sinica)

Email: wdw@iis.sinica.edu.tw

Tel: 886-2-2788-3799 #1729

Fax: 886-2-2782-4814

Change Management Strategies: Transforming a Difficult Implementation into a Successful One

Giussi María Victoria^a, Baum Analía^a, Plazzotta Fernando^a,
Muguerza Pablo^b, González Bernaldo de Quirós Fernán^c

^a Clinical Informatics Office, Ministry of Health, Buenos Aires city, Argentina

^b IT department, Ministry of Health, Buenos Aires city, Argentina

^c Government Health Advisor, Buenos Aires city, Argentina

Abstract

The implementation of health Information Technologies (IT) contributes to improve quality of care and management processes. In spite, evidence shows that the rates of IT adoption are not the expected ones. Since 2004, Public Healthcare System of Buenos Aires city has been implementing a Healthcare Information System with a difficult adoption in clinical setting. In December 2015, the Government made some changes that allowed the implementation of an Electronic Health Record in 20 Primary Care Centers. This paper describes the change management strategies that were designed in order to transform a difficult implementation into a successful one. The combination of timely approach to change management, good governance and specialized human resources were keys to achieve this goal.

Keywords:

Electronic health record, Healthcare system, Change management

Introduction

Implementation of health Information Technologies (IT) contributes to improve quality of care and management processes. Provision of quality information for health care decision making, optimization of resources, contribution to the healthcare system giving timely and equitable care, are some of the benefits that can be achieved. Although IT benefits are clear and well described, evidence shows that the adoption of new information systems is not an easy task, and the utilization rates are not the expected for the current times [1].

There are many issues that can become barriers in the implementation of a health IT project, if they are not timely taken into account. For example: political and economical support to ensure governance, the design of a strategic plan with clear objectives in short, medium and long term, the lack of external regulations and internal processes that provides a legal framework according to the needs, the definition and adoption of interoperability standards, and change management strategies to transform the organizational status quo [2]. Evidence has shown that more than a half of IT projects result in failure [3]. Despite inevitable failures, medical informatics leads to many successes. Implementing information systems into complex healthcare organizations requires an effective blend of technical and organizational skills [4]. Therefore, the technology field offers challenging possibilities and new tools that demand new approaches, without disregarding the human factor [5].

Since 2004, the Public Healthcare System of Buenos Aires city has implemented a Health Information System (HIS) called SIGEHOS (Sistema de Gestión Hospitalaria -Hospital Management System-). SIGEHOS was mainly adopted in hospitals for administrative purposes such as patient identification, manage medical appointments and billing services according to patient insurance. In 2012, SIGEHOS was provided with clinical record capabilities and tried to implement in Primary Care centers; however, all approaches and strategies to accomplish significant results were unsuccessful. The arrival of a new Government in December 2015 brought in different policies on health IT and the possibility of another chance.

This paper describes the new approach and strategies designed after 2015 in order to transform a difficult implementation into a successful one.

Methods

Setting

Health Care Network

Buenos Aires is the capital city of Argentina and its largest and most important one. According to the 2010 census, the city's population is 2,890,151 inhabitants [6].

The Argentine's Public Healthcare System provides free services to its entire population but much different insurance models often coexist on the same patient. According to 2014 statistical data, the average of the population residing in Buenos Aires city is insured only by the public system is 17.8%. This percentage reaches 31.2% in the south of the city, which is the more unprivileged [7].

The healthcare network is formed by 33 Hospitals and 43 Community and Primary Care Centers (Centros de Salud y Acción Comunitaria - CESAC). The city is structured into 12 areas in order to organize health care delivery (Figure 1). In 2015, the Buenos Aires's Public Healthcare System provided approximately 9,600,000 of consultations in ambulatory setting [8].

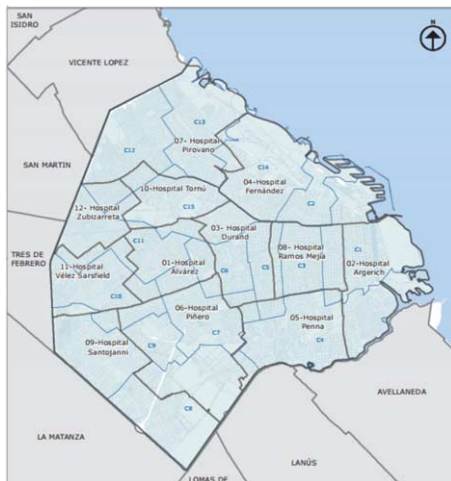


Figure 1 - Geographical distribution of healthcare areas

Software

SIGEHOS is web-based and in-house developed software (Figure 2). From 2004 to 2012, the first version of SIGEHOS was implemented in 20 hospitals, including modules that allowed patient identification, schedules management, hospital admission and billing, pharmacy stock management, registry of corpses and management reports. In 2012, an electronic health record (EHR) was developed and incorporated into SIGEHOS with the aim of replacing paper clinical records and getting the benefits of the electronic format.



Figure 2 - SIGEHOS Interface

First EHR implementation

Since 2012, the Government intended to implement an EHR and develop initiatives to improve the infrastructure of the healthcare institutions.

It was not until 2014 that said implementation started in Primary Care Centers, but by October 2015 it was adopted only in 1 of the 43 CESAC.

The following problems were encountered:

- Difficult negotiations with each CESAC due to lack of governance, and project management were in hands of a technical department with few healthcare professionals participating
- Lack of a multidisciplinary team which understands how to deal with healthcare professional's resistance to change.
- The definition of standards was inappropriate to represent the different domains of clinical knowledge

Study design

This is a descriptive paper about the strategies implemented in the primary healthcare network in Buenos Aires city since December 2015. These strategies were based on a better health care strategic planning and on the redesign of the primary care network [9].

Results

Change management strategies

The strategic plan comprises 7 fields related to Health IT projects [10].

1. Change of project leadership: The Clinical Informatics (CI) Office was created and coordinated by a physician specialized in health informatics. Thirty professionals from different disciplines such as medicine, nursing, health informatics, systems analysis and engineering, sociology, psychology, anthropology, social communication and education sciences, were incorporated to the implementation team (Figure 3). All these professionals set up a great interdisciplinary workforce along with the IT department.

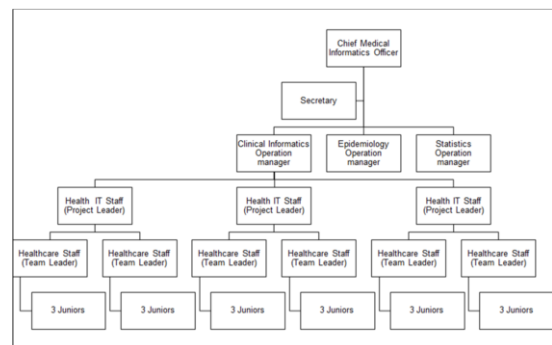


Figure 3 - Organization Chart

2. Organizational Policies and Procedures: The standardization of key processes, such as patient identification, was needed. In this regard, a special area for internal policies and procedures was created with the purpose of redesigning, documenting and publishing processes and workflows along with the implementation of the EHR.
3. Standards for interoperability
 - *Person Identification Services:* The patient identification module was strengthened with rules for unique and unequivocal identification. A Master Patient Index (MPI) with an audit module was developed based on Person Identification Services (PIDS) and CORBAMED standard [11].
 - *Clinical Terminology Services:* Several SIGEHOS EHR modifications were requested to the development team, taking into account physicians complaints and recommendations before extending the implementation. One of these complaints was the ICD-10 codification drop-down menu. Clinical Terminology Services (CTS) based on SNOMED CT as reference vocabulary were included to

automatically coding diagnoses and problems list [12].

4. User Centered Design (UCD): One of the major modification from the first EHR version was the shift to a problem oriented EHR. The new version allowed professionals to record clinical notes, anthropometric values, vital signs and vaccination (Figure 4). The developing team worked with UCD techniques in new functionalities like: family clinical record, pregnancy registry sheet, a computerized physician order entry (CPOE) module, recording and visualizing social, environmental, community and epidemiological data, all of which are keys for Primary Care and Public Health [13].

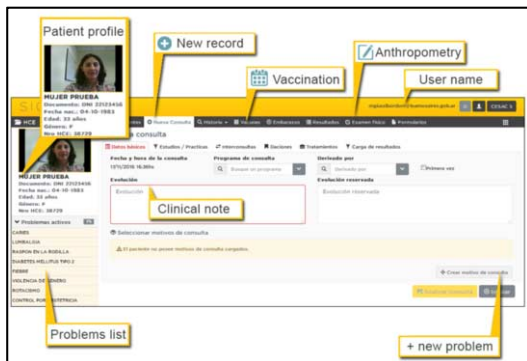


Figure 4 - SIGEHOS EHR Interface

5. Training, Communication and Support

- **Training characteristics:** The implementation team committed to train the final users of the system (software) in situ with a theoretical-practical approach. A team of 4 people performed the training 7 hours a day during a period of 4 weeks.
- **Communication and Support:** A web-based portal for Primary Care was developed; this webpage was the entrance door for access the EHR and its support system. This web page allowed unifying communication between clinical staff and delivering news about health care to final users.

6. Qualified human resources for continuity: In November 2016, a multidisciplinary residency program on health information systems was created. This program is oriented to professionals from different disciplines and based on a similar training program already in practiced in Argentina with more than 10 years of experience [14, 15]. Once residents finish the program, they should be capable of:

- Analyze Health Information Systems (HIS) and detect opportunities for improvement in healthcare organizations.
- Design HIS to accompany healthcare processes and contribute with quality information for continuity of patient care.
- Implement HIS that allows obtaining quality data for the management of public health policies

7. External Rules, Regulations, and Pressures

- **Legal framework:** In October 2016, Buenos Aires legislature passed the EHR law [16]. This law provided an important legal framework to support the implementation of the project and help to overcome the natural resistance to change. It also guaranteed the rules to carry out standardized implementations between public and private health care providers.
- **Specialist Advisory:** To develop the project, the Government received advice from a specialized physician with almost 20 years of healthcare management experience and a pioneer in health information system implementation in the region [17].

Implementation results

Implementation results are described before and after applying the strategies described above as parameters of implementation success.

Goals

At the beginning of the project the goal was to implement the EHR in 14 CESAC by the end of 2016, taking into account that 6 months were needed for setting up the team, gather processes and change the SIGEHOS software. The first implementation was in June 2016, and after applying new strategies the team was able to increase the initial target until 20 CESAC were fully implemented by December 2016.

Users

The adoption of the EHR increased from 27 users in 2015 to 1200 on December 2016, 90% of them was healthcare professionals and 10% administrative staff.

Among the healthcare professionals users, most frequent specialties were pediatrics, nursing, general medicine, psychology, obstetrics, internal medicine and social service (Figure 5).

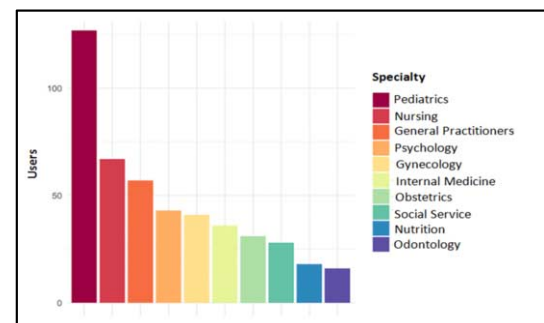


Figure 5 - SIGEHOS users by specialties

Identifying patients

In a similar way, the new approach helped to increase more than 10x times the number of patient registered in the MPI in a 6 months period reaching a total of 52,580 patients by December 2016 (Figure 6).

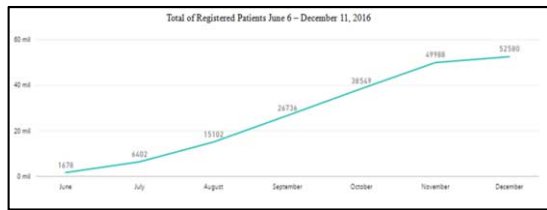


Figure 6 - Total of registered patients

Progress of Clinical Notes

The more professionals used the EHR the more clinical notes were registered, going from 10,000 notes to a total of 75,281 by December 2016 (Figure 7).

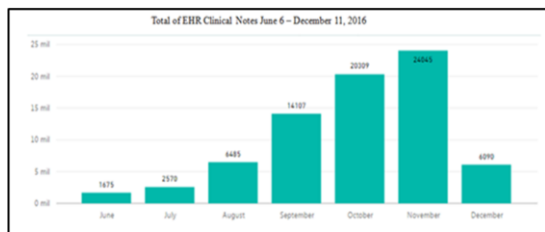


Figure 7 - Total of clinical notes

The distribution of clinical notes by specialty showed that the highest percentage of them was used in pediatrics, followed by gynecology and internal medicine. Nursing occupied the 4th place along with obstetrics (Figure 8).

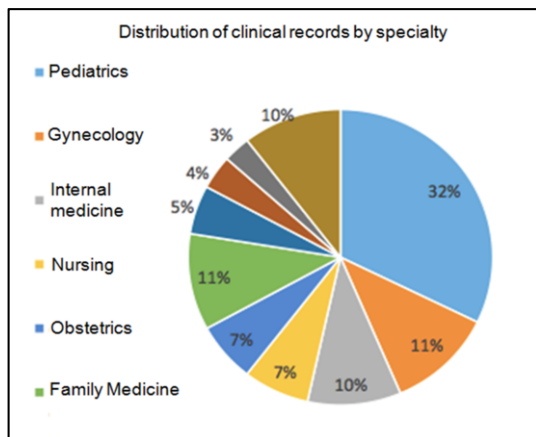


Figure 8 - Clinical records by specialty

Discussion

The governance model implemented since December 2015 allowed the successful implementation of a health information system in Buenos Aires city. The implementation was carried out within the framework of a strategic ministerial aim consisting of strengthening primary care network and achieving an integrated and efficient healthcare system that provides timely, equitable and progressive care to people.

A key issue was the standardization of processes and the incorporation of interoperability standards for unequivocal patient identification. Primary clinical data capture should be as expressive as possible, then semantic coding and control should be transparent to the user. First EHR version considered mandatory the ICD-10 codifying by professionals and it caused resistance. Terminology services were incorporated in second EHR version and users could write free text overcoming barriers. This is one of the most recognized interventions and ensures the scalability of the project [21].

Another key aspect was the careful building and shaping of the implementation team. A blend of IT professionals and clinicians cannot become a team of health informatics specialists. There was a need for specialists in health informatics to be a "bridge" between the two disciplines and to be the guide through the changes of the organizational culture. The incorporation of healthcare professionals specialized in health informatics helped to reduce the time needed to implement the software and overcome preexisting barriers [5]. Moreover, in order to maintain the project beyond any political changes, an educational program was built to train an ongoing qualified workforce.

Healthcare professionals usually have little resistance to change unless the change is beyond their control or is intended to modify existing work conditions. In order to decrease healthcare professional resistance to change it was necessary to consider healthcare workflow in the redesign of processes and the EHR. This approach avoided the misuse of the systems and the poor quality of the recorded information by professionals [4, 22, 23].

Lack of regulations and legislation about eHealth is an important barrier to a HIS adoption. Working together with a group of lawyers, lawmakers and other professionals made it possible to define a legal framework for the implementation of the EHR.

The development and implementation of a robust and scalable health information system involves the detailed knowledge of the business rules of each healthcare organization. Despite the fact that good software solutions can be imported, the strategies of implementation and the "peopleware" are proper of each institution [24].

It is hoped that this document will be useful for similar experiences. It is necessary to continue working on the implementation of this project and to achieve an evaluation through a validated instrument.

Conclusion

Implementing Health Information Systems in Public Health faces a major challenge that requires political good governance, human resources with specific skills and a health information system able to adapt to the culture and organizational context.

This paper describes first months of an EHR implementation in Primary Care in Buenos Aires city. Good governance, a strategic implementation plan and the conformation of a interdisciplinary work team with human resources trained in health informatics, were key components for carrying out a large-scale project and handling adequately the resistance to change.

References

- [1] S. Ajami and T. Bagheri-Tadi, Barriers for Adopting Electronic Health Records (EHRs) by Physicians, *Acta Inform Medica* **21** (2013), 129–134.

- [2] J.K. Pinto and S.J. Mantel, The causes of project failure, *IEEE Trans Eng Manag* **37** (1990), 269–76.
- [3] S. Sharon, More than half of IT projects still failing | CIO, 2016. Available at: <http://www.cio.com/article/3068502/project-management/more-than-half-of-it-projects-still-failing.html> (accessed December 19, 2016).
- [4] N.M. Lorenzi and R.T. Riley, Managing change: an overview, *J Am Med Inform Assoc* **7** (2000), 116–124.
- [5] M.J. Ball and J. Douglas, Human Factors: Changing Systems, Changing Behaviors, 2002.
- [6] Gobierno de la Ciudad de Buenos Aires 2016. Available at: <http://www.buenosaires.gob.ar/laciudad/ciudad>.
- [7] Encuesta Anual de Hogares 2014 Ciudad de Buenos Aires. Síntesis de resultados. 2014.
- [8] Anuario Estadístico 2015 Ciudad de Buenos Aires. 2015.
- [9] M.G. Bordoni, F. Plazzotta, A. Baum, C. Ilc, and B.deQ.F. Gonzalez, Elaboración e implementación de una Agenda Digital en Atención Primaria en la Ciudad Autónoma de Buenos Aires. CBIS 2016.
- [10] D.F. Sittig and H. Singh, A New Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems, *Qual Saf Heal Care* **19** (2010), 68–74.
- [11] Person Identification Service (PIDS) Specification. IDX Syst Corp Copyr IONA Technol PLC Copyr Oacis Healthc Syst Copyr 1997.
- [12] M.L. Gambarte, A.L. Osornio, M. Martinez, G. Reynoso, D. Luna, and F.G.B. de Quiros, A practical approach to advanced terminology services in health information systems, *Stud Health Technol Inform* **129** (2007), 621–625.
- [13] J.S. Ancker, C. Chan, J. Chelico, S. Khan, S. Mortoti, K. Natarajan, and et al, Redesigning electronic health record systems to support public health, *J Biomed Inform* **40** (2007), 398–409.
- [14] B.deQ.F. Gonzalez, D. Luna, P. Otero, A. Baum, and D. Borbolla, Spreading knowledge in medical informatics: the contribution of the hospital Italiano de Buenos Aires *Yearb Med Inform* (2009)147–152.
- [15] A. Baum, F. Plazzotta, D. Canosa, D. Borbolla, P. Otero, D. Luna, and et al, Especialistas en Informática Médica: 10 años de experiencia de un programa de residencias médicas en Sudamérica, INFOLAC 2011, 2011.
- [16] Boletín Oficial de la Ciudad de Buenos Aires Nro 5019 n.d.:388. Available at: <http://boletinoficial.buenosaires.gob.ar/#>.
- [17] D. Luna, F. Plazzotta, C. Otero, A. Baum, S. Benitez, B.deQ.F. Gonzalez, Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires | Publication | Comisión Económica para América Latina y el Caribe. n.d.
- [18] P. Navarro, D. Luna, B.deQ.F. Gonzalez, and Cols, Y. Construcción de un Sistema de Identificación de Personas con énfasis en los procesos de Control de Calidad de los datos, INFOLAC 2008.
- [19] J.S. Rose, B.J. Fisch, W.R. Hogan, B. Levy, P. Marshal, D.R. Thomas, and et al, Common medical terminology comes of age, Part One: Standard language improves healthcare quality, *J Healthc Inf Manag* **15** (2001), 307–318.
- [20] J.S. Rose, B.J. Fisch, W.R. Hogan, B. Levy, P. Marshall, D.R. Thomas, and et al, Common medical terminology comes of age, Part Two: Current code and terminology sets--strengths and weaknesses, *J Healthc Inf Manag* **15** (2001), 319–330.
- [21] S.T. Rosenbloom, R.A. Miller, K.B. Johnson, P.L. Elkin, and S.H. Brown, Interface terminologies: facilitating direct entry of clinical data into electronic health record systems *J Am Med Inform Assoc* **13** (2006), 277–88.
- [22] J.S. Ash, Factors and Forces Affecting EHR System Adoption: Report of a 2004 ACMI Discussion, *J Am Med Informatics Assoc* **12** (2004), 8–12.
- [23] L. Gottlieb, R. Tobey, J. Cantor, D. Hessler, and N.E. Adler, Integrating Social And Medical Data To Improve Population Health: Opportunities And Barriers, *Health Aff* **35** (2016), 2116–2123.
- [24] C. McCarthy and D. Eastman, Change Management Strategies for an Effective EMR Implementation, In: HIMSS, editor. HIMSS, 2010.

Address for correspondence

Giussi Bordoni María Victoria, MD. Ministry of Health, Buenos Aires city. Argentina. mgiussibordoni@buenosaires.gob.ar
 Phone number: +54 9 (011) 4123-3215; Mobile number: +54 9 (2317) 415348

Evaluate the Usability of the Mobile Instant Messaging Software in the Elderly

Tzu-Ning Wen^a, Po-Liang, Cheng^{a,b}, Po-Lun Chang^a

^a Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

^b Buddhist Tzu-Chi Medical Foundation, Dalin Tzu-Chi Hospital, Chiayi, Taiwan, China

Abstract

Purpose: Instant messaging (IM) is one kind of online chat that provides real-time text transmission over the Internet. It becomes one of the popular communication tools. Even it is currently an era of smartphones, it still a great challenge to teach and promote the elderly to use smart phone. Besides, the acceptance of the elderly to use IM remains unknown. This study describes the usability and evaluates the acceptance of the IM in the elderly, who use the smartphone for the first time. This study is a quasi-experimental design study. The study period started from October, 2012 to December, 2013. There were totally 41 elderly recruited in the study. All of them were the first time to use LINE app on the smartphones. The usability was evaluated by using the Technology Acceptance Model which consisted of four constructs: cognitive usability, cognitive ease of use, attitude and willingness to use. Overall, the elderly had the best "attitude" for LINE APP communication software, with the highest rating averaging 4.07 points on four constructs, followed by an average of 4 points on "cognitive usefulness". The scores of "cognitive ease of use" and "willingness to use" scores were equal which are an average score of 3.86. It can be interpreted that (1) the elders thought that the LINE APP as an excellent communication tool for them; (2) they found the software is useful (3) it was convenient for them to communicate. However, it was necessary to additionally assist and explain the certain functions such as the options. It would play a great role in the "willingness to use". The positive acceptance of LINE APP in elderly refer to the probable similar acceptance for them to use other communication software. Encouraging the willingness the elderly to explore more technology products and understanding their behavior will be the basic knowledge to develop further software.

Keywords:

Technology; Instant messaging (IM); Technology Acceptance Model

Introduction

Aging of the population is an irreversible change. The effective promotion of using the technologic product in the elderly can enhance the "active ageing". In the past studies have shown the dependence of an elderly person relied on the

several factors which were recognized as primarily functional and cognitive impairment, chronic diseases, a diminishing social network, and a low level of physical activity. However, effectively using the IM can make their lives more abundant and enhance their interpersonal relationships.

The alleviation of the chronic diseases and enhancement of the subjective happiness result in higher self-appraisal in health condition. By the contrast, it is not prone to develop into depressive mood and the related symptoms.[2]

The effective use of mobile communication software for the elderly will reduce the social isolation, conversely, make them confident and be willing to participate in the community activities and their interests. [3] [4]. Therefore, it is a solution for the most elderly people to get positive attitudes towards technology [5] [6].

In the past, many studies have explored the use of technology acceptance patterns to compare the acceptance of application-related technologies such as Mobile Health, Telehealth, and Telecare for the elderly.

The Technology Acceptance Model (TAM) was developed by Davis (1989) on the basis of rational behavioral theory TRA. This model to predict and explain a person's adoption of information technology. It consists of (1) the usefulness awareness (2) the ease of the use (3) the use of attitude (attitude toward use) (4) the use of the behavioral (actual use) [7]. In TAM, usefulness awareness and ease of use cognition are considered to directly affect the use of attitude [8]. When a technology is easy to use, it is easier for the user to think that it is useful [9] [10].

Therefore, the concept of technology acceptance model was applied to develop the the certain questionnaire which will clarify the relationship between cognitive use, cognitive usefulness, use attitude and willingness to use communication software in the elderly. The feasibility of this model-based study will be the blueprint and reference of the further system construction.

Methods

This study is a quasi-experimental design study. The study period started from October, 2012 to December, 2013. The program was held at a regional urban hospital in Taipei City. Each participant was provided with an android-based smart phone including a charger, a memory card, and a mobile phone number. There are two lectures during the study period.

The first lecture was held to instruct the participants to use the basic function of the mobile phone. The second lecture in one month later was held to instruct the participants to operate the IM app “LINE” and its group chatting. The feedback from the participants was collected in the same lectures. Three months after the first lecture, the participant completed the TAM-based questionnaire. The feedback and trouble analysis were also done.

The reliability and validity of the questionnaire

In order to ensure the validity of the contents of the questionnaire, a seven-experts panel meeting was held to assess the relevance and the completeness. The four experienced experts devoted themselves to geriatric researches and the other three engineering background experts dedicated to the programming on the mobile devices. The experts provided suggestion of the relevance, comprehensiveness of the content and the grammar. The questionnaire contains four construct as the cognitive usefulness, cognitive ease of use, the use of attitude and the use of will. Besides, the troubles while using LINE all were also collected and analyzed. Acceptance were measured on a Likert scale of five points, with 1 being strongly disagreeable, 2 being disagree, 3 being no opinion, 4 being consent, and 5 being very agreeable. The expert evaluates the validity of the scale according to the item, content and description of the questionnaire. Content validity index (CVI) was 0.63 and the internal consistency (Conbach's alpha) was 0.824.

Data processing and analysis

The results were analyzed with SPSS 23 statistical software to describe the distribution of variables. Descriptive t-test and one-way ANOVA were used to analyzed the effects of the four constructs from the individual factors such as gender, age, educational level, the experience of using traditional mobile phones or smartphone and the understanding of LINE APP and other attributes.

Results

There were 49 participants initially recruited, and only 41 participants completed the survey. The (Figure 1).

The basic properties of the study object

There were 61% female and 39% male participants. The majority of age was 60-69 years old (60.90%). There were 34.1% college-educated participants. The majority of the monthly family income was the group less than USD\$940.97 were family income (39%). Most of the participants (68.3%) had the limited experience of using traditional mobile phones and smart phones less than five years. However, 28 of 41 (68.3%) participants tried to download LINE app during the study period. (Table 1)

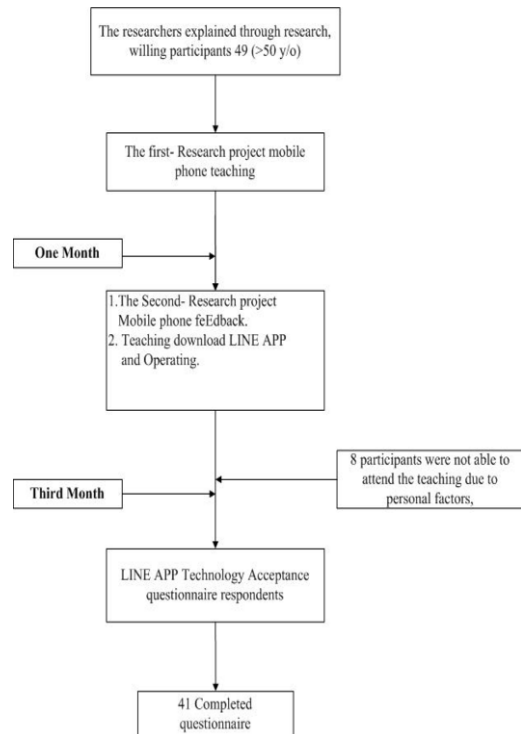


Figure 1 - Study the flow chart

Table 1 - The demographic data of the subjects described the statistical results. (N = 41)

Variable	n	%
Gender		
Male	16	39%
Female	25	61%
Age		
50~59 y	7	17.10%
60~69y	27	65.90%
70~79y	6	14.60%
>80y	1	2.40%
Education		
Illiterate	1	2.40%
Elementary school	2	4.90%
Junior High school	8	19.50%
Senior High school	11	26.80%
College / University	14	34.10%
Graduate school	5	12.20%

Monthly Home Income

< USD\$940.97	16	39.00%
USD\$941~1568.26	3	23.00%
USD\$1568.29~2195.57	10	14.40%
USD\$2195.60~2822.88	10	24.40%
>USD\$2822.92	2	7.90%

Experience of using traditional mobile phones and smartphone experience

< 5 years	28	68.30%
5~10 years	4	17.10%
11~15 years	7	9.30%
>15 years	2	4.90%

The experience of downloading LINE APP

Yes	28	68.30%
No	13	37.10%

The technology acceptance of the LINE app

The questionnaire construct is divided into "cognitive usefulness", "cognitive ease of use", "use attitude" and "use will" totally four major aspects consisting of 17 questions. Overall, the elderly had the best "attitude" for LINE APP communication software, with the highest rating averaging 4.07 points on four constructs, followed by an average of 4 points on "cognitive usefulness". The scores of "cognitive ease of use" and "willingness to use" scores were equal which are an average score of 3.86. It can be interpreted that (1) the elders thought that the LINE APP as an excellent communication tool for them; (2) they found the software is useful (3) it was convenient for them to communicate. However, it was necessary to additionally assist and explain the certain functions such as the options. It would play a great role in the "willingness to use".

I think it is useful to communicate through the LINE APP communication software (4.1), and to use the voice communication function (4) for a score of 4 or above. (4.02), C1 "I think it is wise to use the LINE APP communication software" (4.07), D2 "I think this tool is worth promoting" and D3 "I am willing to Recommend this tool to friends and family to use" average score is 4.05 points. There were six items in the "percentage" of the distribution, with "consent" and "very much agree". More than 60% of the total, respectively, were "using the LINE APP communication software to meet my mobile communication needs" (63.4%). (65.9%), B1 "LINE APP communication software is very easy to operate" (63.4%), B2 "LINE APP to graphic display of each function project, so that the use of operation (65.9%), B4 "Limitation of using LINE APP will be less than traditional SMS, 61% of mobile phone calls", and B6 "Communicate through LINE APP, I think it is simpler than traditional mobile phones" (63.4%). Can explain the majority of the elderly that LINE APP in the use of communication is very

useful for them, and the ease of use on the operation can have a certain understanding. This means that the elders have some acceptable usability for LINE APP acceptance (Table 2).

The impact of different attributes on four dimensions of LINE APP technology acceptance

($t = 2.51$, $p < 0.05$), which indicated that the educational level of LINE APP was significantly different ($P < 0.05$), which indicated that there were significant differences in LINE APP between different educational backgrounds. The willingness of the use of the difference in the higher level of education for the elderly have a higher LINE APP use. Other attributes were not statistically significant (Table 3).

Table 2 - LINE APP Acceptance is used.

	Gender /M±SD		<i>t</i>	<i>p</i>
Variable	Male N=16	Female N=25		
LINE APP usefulness				
A1 Can quickly use voice calls	4.19±0.75	3.88±0.72	1.306	0.19
A2 Can easily, simply, and quickly upload messages	4.06±0.68	4±0.70	0.28	0.78
A3 Use social networking to share your current status	4.13±0.80	3.84±0.62	1.27	0.21
A4 This tool is able to meet my needs for mobile communications applications	4.13±0.71	3.84±0.62	1.34	0.18
A5 LINE APP do communicate I think it is useful	4.31±0.60	3.96±0.53	1.95	0.05*
LINE APP ease to use				
B1 LINE APP is easy to operate	3.81±0.91	4±0.57	-0.73	0.47
B2 LINE APP	4±0.63	3.84±0.55	0.85	0.39
B3 This tool graphs each functional item to make it easier to understand the operation	3.75±0.85	3.88±0.6	-0.57	0.57
B4 Use of this tool is limited to less than traditional text messaging and mobile phone calls	3.94±0.92	4±0.70	-0.24	0.8
B5 Even without operating instructions or human guidance, you can quickly learn to operate this tool	3.44±0.96	3.72±0.73	-1.06	0.29
B6 Communicating through LINE APP is simpler than traditional SMS	4±0.73	3.84±0.62	0.74	0.45
LINE APP Willingness to use				
C1 I think using the LINE APP is a wise choice	4±0.89	4.12±0.52	-0.54	0.59
When using LINE inconvenience for				
D1 LINE APP in communicating with others there is my needs	3.81±0.91	4.04±0.61	-0.95	0.38
D2 I find this tool worthy of promotion	3.88±1.08	4.16±0.62	-1.06	0.29
D3 I would recommend to friends and family use this tool	3.94±1.06	4.12±0.60	-0.7	0.48
D4 If this tool charges, I still willing to use	2.81±0.98	3.64±0.95	-2.6	0.11
D5 I would like to use this APP interactive communication with others	3.88±1.08	3.96±0.67	-3.09	0.75

Table 3 - Effect of Different Structures on LINE APP Acceptability

Variable	LINE APP usefulness		LINE APP ease to use		LINE APP Willingness to use..		When using LINE inconvenience for me....	
	t/F	p	t/F	p	t/F	p	When using LINE inconvenience for me....	p
Gender	1.5	0.14	-0.31	0.75	-0.54	0.59	-1.49	0.14
Age	2.51	0.74	2.21	0.96	0.9	0.44	0.94	0.43
Education	0.6	0.7	0.34	0.88	0.57	0.71	2.51	0.04*
Month In-come(Home)	0.48	0.74	0.53	0.7	0.13	0.96	0.02	0.99
Use of traditional mobile phones and smartphone experience (time)	2.25	0.09	0.72	0.54	0.74	0.53	0.07	0.97
Do you know about LINE APP?	0.52	0.6	0.82	0.56	0.95	0.34	0.86	0.39

The elderly for the use of LINE APP troubled analysis

In the study of elderly people to use LINE APP during the troubled items are three items are "1.when passed chat messages too noisy", "2.For not familiar with, not more convenient to talk about." The first item "When passed chat messages too noisy" average score accounted for a maximum of 34.6 points; followed by the second item "For not familiar with, not More convenient to talk about." Average score 3.41 points; the third item "Chat interface complex, do not know how to click" average score 3.39 points. If you get chat messages too noisy (46.3%), you get the "Consent" and "Strongly Agree" options from a distribution percentage of more than 40%. It can be explained that there are still some elderly people who think that the use of LINE APP for them, we chat when the message may cause their distress, (Table 4).

Table 4 - The use of LINE APP troubles

When using LINE inconvenience for me.... 3.42					
Questions	M±SD	Very disagree	Disagree	General	Agree
1.When passed chat messages too noisy	3.46±0.97	2(4.9%)	4(9.8%)	12(29.3%)	19(46.3%)
2.For not familiar with, not more convenient to talk about.	3.41±0.89	2(4.9%)	1(2.4%)	20(48.8%)	14(34.1%)
3.Chat interface complex, do not know how to click	3.39±0.97	3(7.3%)	1(2.4%)	18(43.9%)	15(36.6%)

Discussion

From this study, the use of the LINE for the elderly to accept the degree is good, not only that this is a useful communication software, for them more feel is an easy to use tool. But in the use of the operation must still be someone to help guide the various buttons between the various functions, so to promote the elderly more convenient operation.

In Davis's (1989) theory of technology acceptance, cognitive usability tends to influence cognitive usability, and both factors directly influence attitudes toward use. In the present study, the elders had the highest ratings on the "attitude" of the

LINE APP communication software, followed by the "cognitive usefulness" and the "ease of use" and the "willingness to use" scores. It can be explained that the elderly in the software that can be used for the software, but also in use that it is useful and easy to use, the relative will be willing to use. The higher the education level of the elderly for LINE APP have a higher use of will. So for the elderly in terms of LINE APP communication software is acceptable.

In addition, this study explores the use of LINE APP communication software for the elderly, showing that the use of communication software chat rooms of the message transmission is a higher problem for them; secondly, if you are unfamiliar with the people in the LINE APP communication software chat room chat is a bit embarrassing situation, do not know what to talk about? Finally, LINE APP communication software button complex, so that the elderly do not know how to press the button. From this it can be seen that if they further instruct the elders to confirm the accuracy of each step of their keys and to confirm their intention for each key, they can self-mute the voice of the message or reduce the sound, To complete. In addition to between the elderly in the LINE APP communication software chat room do not know what to talk about? If researchers are able to deliver a thematic message in this group, leading the elders to the topic, it may be possible for the elders to participate in the discussion.

Conclusions

This study is an experimental study, the technology acceptance questionnaire for data collection, for the elderly > 50-year-old home for the recipient to investigate the results, if you want to apply the results to other age levels of scientific and technological acceptance, has its inference limits. Suggested that the future can do the investigation of different age groups, in-depth understanding of different age groups for the use of communication software habits and recommendations.

References

- [1] C.M. Perissinotto, I. Stijacic Cenzer, K.E. Covinsky, Loneliness in older persons: a predictor of functional decline and death, *Arch. Intern. Med.* 172 (2012) 1078–1083.
- [2] J.Chopik. The Benefits of Social Technology Use Among Older Adults Are Mediated by Reduced Loneliness. *Cyberpsychol Behav Soc Netw.* (2016), 551-556.
- [3] M. Lupp, T. Luck, S. Weyerer, H.-H. König, E. Brähler, S.G. Riedel-Heller, Prediction of institutionalization in the elderly: A systematic review, *Age Ageing* 39 (2010) 31–38.
- [4] Y.R.R.Chen, and P.J.Schulz, The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review. *Journal of medical internet research.* (2016). 18(1) 1-18.
- [5] K. Chen, and Alan H.S. Chan, A review of technology acceptance by older adults, *Gerontechnology* (2011), 1-12.
- [6] T. Sims, A.E. Reed and D.C.Carr. Information and Communication Technology Use Is Related to Higher Well-Being Among the Oldest-Old. *J Gerontol B Psychol Sci Soc Sci.* (2016), pii: gbw130.
- [7] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw, "User acceptance of computer technology: A comparison of two theoretical models," *Management Science.* (1989), 35(8) 982-1002.
- [8] N.M. Suki and N.M. Suki, Exploring the relationship between perceived usefulness, perceived ease of use, perceived enjoyment, attitude and subscribers' intention towards using 3G mobile services. *Journal of Information Technology Management.* (2011), 22(1), 1-7

- [9] P.J. Hu, P.Y.K. Chau, O.R. Liu Sheng, and K.Y. Tam, "Examining the technology acceptance model using physician acceptance of telemedicine technology," *Journal of Management Information Systems*, (1999) 16(2), 91–112.
- [10] M. Igbaria, N. Zinatelli, P. Cragg, and A. Cavaye, "Personal computing acceptance factors in small firms: a structural equation model," *MIS Quarterly*, (1997), 21(3) 279-302.

Augmented Reality in Nursing: Designing a Framework for a Technology Assessment

Hanna Wüller^a, Marcus Garthaus^a, Hartmut Remmers^a

^a School of Human Sciences, Osnabrück University, Osnabrück, Lower Saxony, Germany

Abstract

During the last decade, research emphasized the wide range of possibilities of augmented reality (AR). At the same time, Information Technology usage in nursing increased. The question occurs if AR can have reasonable fields of application in nursing. Nursing hinges strongly on the emotional and physical relationship between patients and their nurses. This may lead to ethical conflicts while implementing AR leading to special challenges. Therefore the realization of a technology assessment (TA) seems to be reasonable. We designed a framework for a TA of AR in nursing through workshops with nursing scientists and practical partners. The framework is designed to address ethical aspects of AR in nursing through techno-ethical scenarios, a plausibility assessment, and a participatory approach.

Keywords:

Nursing; Technology Assessment; Ethics

Introduction

Usage of Information Technology (IT) in nursing is constantly increasing [1]. While technology in nursing may have some positive effects on nursing processes, the special characteristics of nursing should be considered when designing technology for nurses. These characteristics are: high emotionality, physical contact, and the importance of the relationship between the patients and their nurses. Within these characteristics the nurse has to take the specific situation of the patient into account and act on an individual basis [2].

The special characterization of nursing, the nursing environment, and the developments and trends in nursing technology shall be taken into account when designing the framework for Technology Assessment (TA) in nursing.

During the past decade technological advancements in augmented reality (AR) have been significant. Thus, AR received a lot of attention, e.g. in the field of architecture, construction, and engineering [3]. No consistent definition of AR exists [4; 5], however, the definition by Azuma et al. is most established [6]. Azuma defines AR as any system that has the following three characteristics:

1. Combines real and virtual
2. Is interactive in real time
3. Is registered in three dimensions

Milgram and Kushino described the virtuality continuum which reaches from the real environment to the virtual environment. AR as part of mixed reality is placed closer to the real environment, whereas augmented virtuality is closer to the virtual environment [7]. Although that definition is

conceived as too generic [4], for the goal of our approach we used it because it includes the widest range of possible applications.

Both definitions share a common concept, they are independent of any technology. They do not refer to special input and output methods as well as tracking procedures. Research in the field of healthcare focuses on the usage of smart glasses without using the term “augmented reality” [8; 9]. In this paper, possibilities of augmented reality are taken into account without limitation to any device such as smart glasses. Whereas, a broad field of prototyped applications exists and a high potential is assumed in different areas. In healthcare most of the implemented use cases are in the field of surgery [10]. For example, anatomic information can be visualized during surgery without interrupting the procedure [9].

Nursing AR research focuses on implementing and evaluating prototypes. For example the implementation and evaluation of a system for hands-free image capture [8] or a system for organizing tasks and monitoring vital signs [11]. These and other studies focus on specific applications without considering the future perspective and ethical aspects. To our knowledge no research on AR in nursing in general exists.

As part of this project, the possibilities and limitations of AR in nursing will be examined. This paper presents a framework for realizing a TA. TAs take intended and unintended consequences of technology into account [12]. One of the challenges of realizing a TA with the aim of influencing technology development is to find the right starting point. If you start the TA in the early stage of an emerging technology, the shape of the technology is very unspecific. In contrast, a TA on a completed technology offers no influence on the technology. The challenge to find the right time to start the assessment is called the Collingridge dilemma [12; 13]. Whereas von Gleich claims that this should not be a problem, he argues that it is most important to find a good way to deal with the lack of information [14]. One way to deal with this lack of information is the usage of appropriate methods.

For conducting a TA, multiple methods may be appropriate. As an example, Grunwald describes two ways just to get prospective results. One way is the exploration of trends and the other is the model based simulation [12]. Djanatliev and German describe a model based simulation supporting prospective healthcare decision making by using an agent-based simulation [15]. Exploration of trends as well as model based simulation rely on available data and allow a quantitative perspective on future trends. For the case of AR in nursing, only prototypes and short-term data exist [16]. Using applications in real life is necessary to produce data, which can be used in model based simulations. Therefore these methods cannot be used.

In contrast, the construction of scenarios allows for the exploration of ethical concerns without the need of previous data. Furthermore, with this approach certain patterns of moral argumentation about new and emerging technology may be included, which are missed in other types of TAs. These aspects are called ethics of new & emerging science and technology (NEST-ethics) [17]. In addition, there is a mutual change of technology and society [18], which Boenink et al. includes in their framework of techno-ethical scenarios [19]. This framework will be described later on.

In the construction of scenarios, one challenge is to assess the plausibility of the technological change. To address this issue Lucivero constructed a plausibility assessment [20]. The single step of the plausibility assessment are described in the construction of our framework. First, exploratory results for the construction of the framework were figured out in workshops. Based on these workshops, we designed a framework for a TA for AR in nursing to address the wide range of possible consequences of using TA in nursing.

Parts of the framework are techno-moral scenarios. The advantage of scenarios is the possibility to take morality change into account [19]. In combination with a plausibility assessment and a participatory approach, these scenarios serve as starting points for future research.

The paper contains four sections: Section one describes the design and the results of the workshops. Section two presents the framework for a TA of AR in nursing. Section three discusses limitations and possible next steps. Finally, section four summarizes the results.

Methods

As shown, many approaches for TAs already exist. All these approaches take different aspects into account. In order to get insights to which kind of TA is needed for AR in nursing, three workshops were designed. As part of the workshops, use cases for AR in nursing were developed. Designing use cases allows us to estimate expectations according to the usage of AR in nursing.

As an introduction to a diversity of perspectives, the workshops had different participants, scientists (workshop I) and practical partners (workshop II and workshop III). All three workshops started with the explanation of AR. The participants and the use cases are described in detail in the following sections.

Workshop I - Scientific Approach

Workshop participants included researchers in the field of nursing and related scientists skilled in nursing practice (n=8). The first part of this workshop was designing a nursing process landscape. The second part discussed use cases in the context of the most relevant nursing processes.

The researchers discussed the characteristics and special aspects of nursing, and confirmed the results of the literature review. They emphasized communication, emotional, physical contact, and priorities of nursing being dependent on the setting. In addition, they identified ethical aspects and privacy concerns as possible obstacles to AR in nursing. Furthermore, hygiene was identified to be an important aspect to take into account when developing an AR application.

Wound care management, food and liquid management as well as drug management were identified as processes with the highest potential for AR in nursing. Furthermore, a

possible reminder function was seen as an advantage. In addition, the possibility of getting legal security was questioned.

Possible use cases of AR in wound care management and drug management will be described later on, as they were part of the other workshops as well. For food and liquid management, AR systems should take into account special diets, give advice if a patient has some intolerance, and measure how much has already been eaten and drunk.

Workshop II and III – Practical Approach

The participants of workshop II (n=7) and III (n=5) had a variety of different backgrounds in the field of nursing (nursing director, IT specialist in a nursing home and operative nurses), IT, and science. The nurses' workplaces involved a nursing home as well as an ambulatory care unit.

In the two workshops of the practical approach, use cases were designed based on the experiences of the participants as well as on advice of the IT experts.

The workshops identified potential processes where AR could offer support. Processes included documentation, contact with experts, physicians and patients' relatives, drug management, handover for the next shift, wound care management, inclusion of new residents, training new employees, reminder function, medical prescriptions, knowledge transfer and care sequences. We further describe the use cases, shift handover, wound care management, and drug management.

Supporting the process of the shift handover is especially relevant for nurses working in ambulatory care units. There is no personal handover as nurses from different shifts do not meet on a daily basis.

Currently, the nurses write down relevant information for the next shift into a book, which does not belong to the regular documentation system. This book should be replaced by a tablet. In addition, the important information should be readable on smart glasses. One of the advantages of this application is the organization of the entries. It should be possible to view entries of one specific shift as well as entries according to one patient. In addition, it should be possible to search within the entries. Furthermore, the tablet offers the possibility of a speech-to-text transcription and of taking pictures. The application on the glasses should allow observing relevant information in the field of view. Which entries are relevant has to be marked at the time of making the entry. Another feature of the digitalized version for the handover should be the possibility of making entries, which are relevant in the future and appear as an entry for the relevant day.

For wound care management, the first documentation of the wound requires nurses to describe the reason for the wound, the kind of wound, the location of the wound, additional actions, and information for other professionals regarding relevant patient data. In addition, the progression of the wound care has to be documented. Relevant aspects are length, width, depth, etc. In addition, it is necessary to take a picture of the wound. With the help of the wound management application it could be possible to take care of the patient and perform wound documentation at the same time.

Furthermore, possible features of an AR application for wound care management could be the serving of therapy proposals and the availability of these proposals at the time of care. In addition, depth and length of the wound should be calculated automatically and be compared to earlier documentation.

The third use case is drug management. AR could assist nurses in preparing drugs for every patient by showing a picture of the correct sample of medicine for every patient. In addition, the system could offer information about where to find the relevant drug. It would no longer be necessary to change view for medication plans while searching for the drug because all relevant information could be provided directly.

Results

The described use cases are diverse. Applications like shift handover, wound care management, and drug management would be needed in specific situations. In contrast, applications to support the whole process of documentation as well as applications for reminder functions need to be used the whole working day.

To examine the relationship between type of application and consequences, a good assessment is needed. According to Halloran et al. ethical aspects will be expressed as soon as stakeholders are working with prototypes and models [21]. According to the wide range of possible applications for AR in nursing using prototypes and models to assess all of them does not seem to be reasonable. As a result this approach is not preferred here.

However, ethical aspects are important for integrating AR in nursing, especially because of the special features of this field, and were explicitly mentioned in workshop I but were not discussed further. In addition, the workshops showed how difficult an imagination of AR technologies can be.

In order to take ethical aspects into account, a framework for a TA to AR in nursing was designed. As the research on AR in nursing is just in its beginnings and no data according to usage of AR in nursing is known, it is necessary to start the research with a prospective method. To shape the future development of the technology it is necessary to conduct a TA in an early stage of the process.

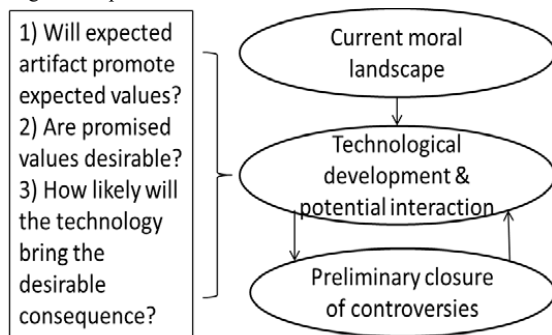


Figure 1 -Framework for a TA to AR in nursing

This framework is a consolidation of three approaches. These are the framework for constructing techno-moral scenarios [19], the plausibility assessment [20] and a participatory approach [12]. Whereas Boenink et al. [19] describe the approach to design techno-moral scenarios in three steps, we integrated a plausibility assessment into the second step and claimed that the development of the scenarios should take place in participatory workshops.

Techno-Moral Scenarios

As previously noted, TAs often miss ethical aspects. One method to address ethical aspects in an emerging technology

are scenarios [22; 23]. As a prospective method, using scenarios allows addressing ethical aspects at an early stage of the development.

To assess ethical aspects of a nascent technology, techno-ethical scenarios are a useful tool because they allow the inclusion of morality change into the considerations [20; 22]. Scenarios can lead to a more reflexive co-evolutionary process and create openings for responsible innovation [22]. Furthermore, scenarios allow reflection on long-term-changes. This is important because “morality normally evolves on the long term” [19].

Boenink et al. describe a framework for developing techno-moral scenarios [19]. Their special focus is on the interaction between technology and morality. In contrast to other scenarios and ethical TAs, their framework takes this interaction into account [19]. The framework consists of three steps. Firstly, the current moral landscape is identified. Secondly, the technological development and its potential interaction with the moral landscape are introduced. Thirdly, a preliminary closure of the controversies based on a historical and sociological analysis is constructed. According to requirements and to evolve the scenarios further step two and three can be repeated [19].

The scenarios describe possible usages of AR in nursing. One possible controversy, which was discussed in the conducted workshops, is the usage of AR in direct contact with the patient. Whereas AR may have the highest usage in this situation, in the current moral landscape the privacy of the patient would have to be protected and AR cannot be used to its full potential. To close this controversy technological developments, which allow data protection or a change of the value of privacy, has to be considered.

Plausibility Assessment

Detailed research is recommended for the construction of scenarios to be able to design plausible scenarios [20; 22]. In addition to the plausibility Rip and Te Kulve consider an informed moderator to be important for being able to intervene in the discussion and offer reasonable arguments [22]. In contrast, Lucivero focuses on plausibility and designed a plausibility assessment. Lucivero argues that following the plausibility assessment allows a grounded construction of scenarios [20]. Scenarios are controlled or grounded through empirical research, they also have a speculative aspect as they explore a potential future in order to trigger imagination [20].

The plausibility assessment consists of three steps. According to Lucivero these steps can be summarized in the following questions:

“How likely is it that the expected artifact will promote the expected values?

To what extend are the promised values desirable for society?

How likely is it that a technology will instrumentally bring about a desirable consequence?” [20]

To answer these questions, relevant data has to be collected. The methods of collecting data in this case are described later on. As one important point Lucivero recommends to describe the technology in detail to reject implausible scenarios and describe two possibilities where appropriate [20].

Participatory Approach

Defining the development of a scenario includes a description of the data resources [24]. In this case, the scenario should be

developed during a workshop. For the preparation of the workshop results from the earlier workshops can be used. Based on these workshops in combination with the literature an overview about the current moral landscape will be handed to the participants. The first step of the workshop is to develop a technological application and potential interaction with it.

To yield ideal results, a participatory approach is used. Nursing IT experts, nursing scientists, ethical experts, elderly people/patients, relatives, nurses and nursing directors shall have the necessary experience for this assessment. In this stage ethical controversies can be detected and discussed. Possible closures have to be included. If AR is able to promote expected values, if these values are desirable, and how likely the technology will bring the desired consequence needs to be discussed when the last controversies could be solved.

The design of scenarios in workshops is a participatory approach which allows the inclusion of stakeholders in the process. For responsible innovation this approach is recommended [25]. Furthermore, a participatory approach is recommended for requirements engineering in general [26].

Discussion

This paper presents a framework for a TA in nursing with the goal of assessing intended and unintended consequences of using AR in nursing. In contrast to existing approaches to ethical aspects in software engineering, this framework shall consider changes in morality and take the plausibility of technological developments into account.

Our empirical results show the necessity of conducting a TA for using AR in nursing. One of the main results was the diversity of imaginable use cases. The early stage of the considerations make it hard to examine ethical controversies. By using a participatory approach in the constructed framework we follow a widely approved method.

Implications for Research and Practice

Due to the growing possibilities of AR and the actual challenges in the field of nursing, the application of AR technologies in nursing seems to be reasonable. The constructed framework provides a starting point to assess intended and unintended consequences of using AR in nursing. The results can be used to shape AR technology in nursing and may serve as a starting point for further research.

The development of AR nursing applications is just in its beginnings. Conducting a TA with the described framework at an early stage of the process may influence the development of the technology positively. Constructed scenarios may be useful for everyone who wants to develop AR applications for the field of nursing. Furthermore, they may lead to interesting results for continued development of AR in other areas.

Directions for Future Research

As a next step the framework has to be implemented and evaluated. An improvement of the framework as well as an examination of ethical controversies shall be the result. Furthermore, this framework may be used to evaluate AR applications in other fields and with other emerging technologies.

The framework has no specific elements for the domain of augmented reality or nursing. Other researchers can therefore adapt and use it for any software development if the

development of the software is in an early stage and the field of application has a high ethical relevance.

Limitations

As the conducted workshops were not evaluated systematically, the presented results have limited impact. Furthermore, due to the small number of participants in the workshops generalization may be limited. Nevertheless, the presented workshops are a first exploratory approach to get knowledge about the complex field of nursing and AR. As nursing and technology follow different logics, the usage of technology may influence the work of nurses negatively [2]. Hülksen-Giesler considers objective information as an obstacle to perceive subjective recognized data which is claimed to be very important to take the specific situation of the patient into account [27]. Whereas this aspect was not directly mentioned in the empirical setting, it was pointed out from the participants with describing communication, emotionality and physical contact as important aspects of nursing. The unspoken question was if technology can support nursing in these aspects.

Furthermore, the proposed TA may require additional methodological steps in order to be applicable to AR that has undergone further development beyond the early stages. This method may be the construction of vignettes based on the scenarios. In contrast to scenarios, vignettes describe not a story but a picture of one possible situation. Vignettes may enhance the discussion about a wide range of ethical concerns. They are more useful if the usage of the technology is already more defined [20]. The vignettes could again be discussed in workshops. The additional usage of vignettes would meet the statement of the Association of German Engineers that a good TA consists of different methods which are used together [28].

Conclusions

The conducted workshops show a need for TA of AR in nursing. The complex topic addresses the need of a framework for a TA which has to take different aspects into account. These are the techno-moral development, the plausibility of expected advantages and disadvantages and the participation of stakeholders. As a next step the TA framework should be evaluated.

References

- [1] U. Hübner, J. Liebe, J. Hüters, J. Thye, N. Egbert, W. Hackl, and E. Ammenwerth, IT-Report Gesundheitswesen-Schwerpunkt: Pflege im Informationszeitalter. , in: *Schriftenreihe der Hochschule Osnabrück*, Hannover, 2015.
- [2] H. Remmers, Natürlichkeit und Künstlichkeit - Zur Analyse und Bewertung von Technik in der Pflege des Menschen, *Technikfolgenabschätzung - Theorie und Praxis* 24 (2015), 11-20.
- [3] X. Wang, M.J. Kim, P.E. Love, and S.-C. Kang, Augmented Reality in built environment: Classification and implications for future research, *Automation in Construction* 32 (2013), 1-13.
- [4] K. Lenzen, M. Park, and B. Kausch, Augmented-Reality-Technologie. Vorschlag einer prozessorientierten Beschreibung, in: *Technology Assessment in der Weltgesellschaft*, Nomos Verlagsgesellschaft mbH & Co. KG, 2007, pp. 455-460.
- [5] W. Broll, Augmentierte Realität, in: *Virtual und Augmented Reality (VR/AR)*, Springer, 2013, pp. 241-294.
- [6] R.T. Azuma, A survey of augmented reality, *Presence: Teleoperators and virtual environments* 6 (1997), 355-385.
- [7] P. Milgram and F. Kishino, A taxonomy of mixed reality visual displays, *IEICE TRANSACTIONS ON Information and Systems* 77 (1994), 1321-1329.
- [8] G. Aldaz, L.A. Shluzas, D. Pickham, O. Eris, J. Sadler, S. Joshi, and L. Leifer, Hands-free image capture, data tagging and transfer using

- Google Glass: a pilot study for improved wound care management, *PLOS ONE* **10** (2015), e0121179.
- [9] S. Mitrasinovic, E. Camacho, N. Trivedi, J. Logan, C. Campbell, R. Zilinyi, B. Lieber, E. Bruce, B. Taylor, and D. Martineau, Clinical and surgical applications of smart glasses, *Technology and Health Care* **23** (2015), 381-401.
 - [10] D. Van Krevelen and R. Poelman, A survey of augmented reality technologies, applications and limitations, *International Journal of Virtual Reality* **9** (2010), 1-19.
 - [11] M. Bang, K. Solnevik, and H. Eriksson, The Nurse Watch: Design and Evaluation of a Smart Watch Application with Vital Sign Monitoring and Checklist Reminders, *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium* **2015** (2015), 314-319.
 - [12] sigma, 2010.
 - [13] D. Collingridge, *The social control of technology*, (1982).
 - [14] A. von Gleich, Prospektive Technikbewertung und Technikgestaltung zur Umsetzung des Vorsorgeprinzips, in: *Konzepte und Verfahren der Technikfolgenabschätzung*, G. Simonis, ed., Springer, 2013, pp. 51-73.
 - [15] A. Djanatliev and R. German, Prospective healthcare decision-making by combined system dynamics, discrete-event and agent-based simulation, in: *2013 Winter Simulations Conference (WSC)*, IEEE, 2013, pp. 270-281.
 - [16] H. Wüller, J. Behrens, M. Garthaus, and H. Remmers, Technologiebasierte Unterstützungssysteme für die Pflege – Eine Übersicht zu Augmented Reality und Implikationen für die künftige Pflegearbeit, in: *ENI UMIT Hall*, 2017, pp. 125-126.
 - [17] T. Swierstra and A. Rip, Nano-ethics as NEST-ethics: patterns of moral argumentation about new and emerging science and technology, *Nanoethics* **1** (2007), 3-20.
 - [18] U. Dolata, *Wandel durch Technik: Eine Theorie soziotechnischer Transformation*, Campus Verlag, 2011.
 - [19] M. Boenink, T. Swierstra, and D. Stemmerding, Anticipating the interaction between technology and morality: A scenario study of experimenting with humans in bionanotechnology, *Studies in Ethics, Law, and Technology* **4** (2010).
 - [20] F. Lucivero, *Ethical Assessments of Emerging Technologies: Appraising the moral plausibility of technological visions*, Springer, 2015.
 - [21] J. Halloran, E. Hornecker, M. Stringer, E. Harris, and G. Fitzpatrick, The value of values: Resourcing co-design of ubiquitous computing, *CoDesign* **5** (2009), 245-273.
 - [22] A. Rip and H. Te Kulve, Constructive technology assessment and socio-technical scenarios, in: *Presenting Futures*, Springer, 2008, pp. 49-70.
 - [23] E.J. Smadja, Scenario Planning as a Tool for Choosing more Ethical Futures: a Case Study in Nanoscale Science and Technology, in: *2006 IEEE Systems and Information Engineering Design Symposium*, IEEE, 2006, pp. 275-278.
 - [24] P.W. Van Notten, J. Rotmans, M.B. Van Asselt, and D.S. Rothman, An updated scenario typology, *Futures* **35** (2003), 423-443.
 - [25] A. Grunwald, Responsible innovation: bringing together technology assessment, applied ethics, and STS research, *Enterprise and Work Innovation Studies* **31** (2011), 10-19.
 - [26] C. Ebert, *Systematisches Requirements Engineering: Anforderungen ermitteln, dokumentieren, analysieren und verwalten*, dpunkt. verlag, 2012.
 - [27] M. Hülsken-Giesler, Der Zugang zum Anderen, *Zur theoretischen Rekonstruktion von Professionalisierungsstrategien pflegerischen Handelns im Spannungsfeld von Mimesis und Maschinenlogik. Osnabrück* (2008).
 - [28] Fachbeirat Gesellschaft und Technik, VDI 3780. Technikbewertung - Begriffe und Grundlagen, in: *2000-09*, VDI, ed., 2000, pp. 334-363.

Address for correspondence

Hanna Wüller, Osnabrück University, 49076 Osnabrück, Germany, hanna.wueller@uni-osnabrueck.de

Design a Learning-Oriented Fall Event Reporting System Based on Kirkpatrick Model

Sicheng Zhou, Hong Kang, Yang Gong

School of Biomedical Informatics, The University of Texas Health Science Center, Houston, Texas, USA

Abstract

Patient fall has been a severe problem in healthcare facilities around the world due to its prevalence and cost. Routine fall prevention training programs are not as effective as expected. Using event reporting systems is the trend for reducing patient safety events such as falls, although some limitations of the systems exist at current stage. We summarized these limitations through literature review, and developed an improved web-based fall event reporting system. The Kirkpatrick model, widely used in the business area for training program evaluation, has been integrated during the design of our system. Different from traditional event reporting systems that only collect and store the reports, our system automatically annotates and analyzes the reported events, and provides users with timely knowledge support specific to the reported event. The paper illustrates the design of our system and how its features are intended to reduce patient falls by learning from previous errors.

Keywords:

Incidental Findings; Accidental Falls; Patient Safety

Introduction

Patient fall has been a prevalent and significant problem in healthcare facilities across the world. The fall rate in acute-care hospitals is between 1.3 and 8.9, mainly ranging from 3-5 per 1,000 occupied bed days [1]. Fall risks are pervasive that not only elder people are vulnerable to falls, but any patient can be at risk for falls due to medical conditions or procedures, such as taking various types of medications, surgeries, or diagnostic tests that can make the patients weak or unstable [2]. A fall with injury adds in average 6.3 days to the hospital stay and costs around \$14,000, which is a huge waste of time and money for both patients and healthcare facilities [3; 4].

To improve this situation, a variety of toolkits for fall prevention have been developed by major agencies and institutes, such as the Agency for Healthcare Research and Quality (AHRQ), Institute for Healthcare Improvement (IHI), Emergency Care Research Institute (ECRI), European Innovation Partnership on Active and Healthy Ageing (EIP-AHA), Institute for Clinical Systems Improvement (ICSI), the Joint Commission, and VA National Center for Patient Safety [2]. Even though these comprehensive toolkits provide systematic introductions and guidelines for building fall prevention programs, a huge gap between these guidelines and practice still exists. First, the actionable solutions are not always applicable and indicative to the conditions where they should apply. This is especially true for less experienced staff. Second, evaluation strategies are inadequately introduced in these toolkits. Third, the compliance over time as directed in the toolkits may gradually reduce. Reinforcement for learning may compete with other clinical priorities and staff often have

difficulties in recalling toolkit contents. Efforts have been made on the fall prevention training programs, yet the expected outcome for reducing patient fall rate remains difficult to achieve[2]. Hereby, how to execute, supervise, and evaluate the fall prevention program in the healthcare settings becomes a major challenge in the patient safety community.

In 1999, the Institute of Medicine's seminal report, *To Err is Human* [5] specifically recommended to establish the event reporting systems within healthcare organizations. Although numerous event reporting systems have been developed since the IOM's report, most of the systems just function as a primary data repository of the reported events. The evidence is very limited to show that event reporting can improve patient safety, and how much influence it can make remains unclear[6]. Nonetheless, learning from errors is still an intuitive way to avoid the recurrence of errors. The reason is that root cause analysis (RCA) can be carried out only when an occurrence of safety event has been reported in details, based on which further actions of improvement will become possible. The event reporting system has great potential to manage the fall cases and provide training for clinicians in long term. Studies on event reporting systems have been limited and fragmented. Most of them were based on the qualitative studies[7-9]. Thus, it has been vague regarding how to build an effective event reporting system and to overcome the technical barriers, the informatics research questions for researchers to explore. The barriers identified are worth an extensive discussion so that timely knowledge support and reporting motivation could be enhanced. In addition, it has been unclear in terms of effective long-term evaluation strategies based on the event reporting systems, triggering the uncertainty about the real effect of the systems. To mitigate or toward removing these issues, the goal of our project is to design a set of learning-oriented and user-centered features to enhance reporting motivation and introduced the Kirkpatrick model into our system toward effective long-term evaluation.

The Kirkpatrick Model

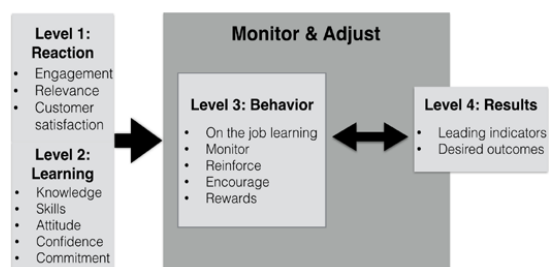


Figure 1 - The Kirkpatrick model [10]

The Kirkpatrick model, is frequently used for training and performance evaluation in various areas, such as business companies, universities and government agencies [11]. It is

reported that the Kirkpatrick model provides the technique for appraisal of the evidence for any reported training program and could be used to evaluate whether a training program meets the expected outcomes of both organizations and the staff [12]. Figure 1 explains an overall structure of the model. Basically, it has four levels, and each level addresses the sub-goals that are necessary to achieve. The four levels will be explained in following sessions in conjunction with the designed features of our system. The model presents the training program in a way that enables the participants not only to learn what they need to know but also to react favorably to the program [13]. No matter in the business area or in the healthcare setting, the training activities have commonality regarding their core challenge, which is how to reach the expected outcomes through improving the staff's learning effect and application of what they have learned in daily work. Thus, in our case, the model has great potential to improve the patient safety event reporting and prevention systems.

The paper summarizes current barriers of event reporting systems, and describes how the Kirkpatrick model serves as a framework in developing a learning-oriented fall event reporting system. Important features including how the system provides instant and specific feedback to the clinicians and builds a reporting and shared-learning culture in healthcare facilities are introduced. The system holds promise in overcoming current barriers of event reporting systems and improving patient safety.

Methods

Identifying barriers of patient safety event reports in healthcare literature

Seven MeSH terms including 'incident', 'reporting', 'system', 'feedback', 'patient safety', 'barrier' and 'challenge' were identified and used to search in the PubMed database with Boolean algebra. Papers published prior to 2004, and not discussing the barriers or challenges of patient safety event reporting systems were excluded. The challenges, barriers and defects of current event reporting systems mentioned in the selected papers were extracted, classified and summarized by two reviewers separately. Then the discrepancies were solved by group discussion, the final results are shown in Table 1.

Building a knowledge base for reporting system

A key component in developing the system is a knowledge base that can provide timely knowledge support to clinicians. The knowledge base contains solutions and classic fall cases. Solutions were extracted and summarized from the patient falls prevention guidelines published by authoritative institutes, including AHRQ, Joint Committee, VA National Center for Patient Safety, Pennsylvania Patient Safety Authority, IHI, EIP-AHA, ICSI and ECRI [2]. Classic fall cases with experts' comments were extracted from AHRQ WebM&M [14], an authoritative peer-reviewed online journal and forum related to patient safety that provides expert analysis of all kinds anonymously reported medical errors cases. The extracted contents were reviewed and modified by domain experts in a Patient Safety Organization (PSO) institute[15].

Applying the Kirkpatrick model in system design

As shown in Figure 1, the model has four levels: reaction, learning, behavior and results. Each level has its own goals that guide the event reporting system design. We designed a set of learning-oriented and user-centered features for the reporting system that aim to achieve all the goals in each level. The developed features will be elaborated in results.

Developing the structure of the fall event reporting system

Our web-based reporting system is mainly comprised of four modules: 1) Data collection module. The AHRQ Common Formats [16] is used as the template for event reporting. 2) A knowledge base used for providing feedback to the users. 3) Data repository module that stores the reported cases. 4) Data analysis module, a key for RCA and long-term evaluation of the outcomes.

Results

Current barriers of patient safety event reporting systems

Eight literature regarding the barriers, challenges and benefits for the patient safety event reporting systems were identified and reviewed. The main barriers of current reporting systems and their occurring frequencies in the literature are summarized in Table 1. All the findings are based on the qualitative studies among front-line staff in healthcare facilities. The barriers can be roughly divided into three types. First and the most important type is that the reporters' doubt about whether the reporting systems can improve patient safety in real practice. Typical examples include "lack of effective feedback" and "no follow up", which indicate the issue of systems usefulness. The second type of barrier is the designing flaws of the systems, such as "poorly designed reporting form", "lack of necessary fields in the system", which make reporters feel unsatisfactory during reporting processes. The third is that cultural factors or working atmospheres in many healthcare facilities do not encourage event reporting. It can be caused by the heavy workload or concerns of taking additional responsibilities. These barriers, all together, negatively impact the utilization and expected outcomes of the reporting systems. Thus, an enhancement in design and implementation of the systems is urgently needed.

Table 1 – Barriers of current patient safety event reporting systems summarized from literature (sorted by frequency)

Barriers	Frequency
Lack of effective feedback [7; 8; 17-21]	7
Fear of reprisals [7; 17; 20; 21]	4
Poorly designed reporting form/system [7; 17; 18]	3
Too busy to report [7; 17; 21]	3
Lack of confidence about the reporting system [17; 21; 22]	3
Unsatisfactory with reporting processes [7; 21; 22]	3
Cultural norms against reporting [7; 8; 22]	3
Lack of systematic analysis of the reports [8; 19; 20]	3
Lack of visible action being taken or a change made after reporting [8; 19]	2
Lack of understanding what to report [17]	1
Lack of perceived value during reporting process [7]	1

Growing a knowledge base for the event reporting system

The knowledge base developed in the system by far has a collection of 137 specific solutions, 45 general solutions, and 29 classic cases for fall prevention. The specific solutions are provided to users when reported events matching the categories of Common Formats. General solutions are provided to reporters as general principles and basic learning materials regardless of the event categories. In addition, classic fall cases with experts' comments from WebM&M are provided as supplementary learning materials for reporters.

Overall design of the reporting system based on Kirkpatrick model

To overcome the barriers of patient safety event reporting systems and evaluate the effect, our system has integrated the following features in correspondence to every level of the Kirkpatrick model.

Level 1: Reaction

Reaction is defined as the degree to which participants react favorably to the learning event [10]. Whether a training program is favorable and engaging is mainly based on to what degree the staff like the training program and are actively involved with the learning experience. Relevance measures whether the participants will have the opportunity to use or apply what they have learned in training on the job [10].

How the system can be favorable, engaging and relevant were carefully considered during the design of the system. We put the concept of user-centered design into first place, which means what users may need and want from the system were always prioritized at all stages of the design process, thus users may use the product as intended and with minimum effort to learn how to use it. The targeted users, including nurses and staff from our institute and our collaborative PSO were involved in the design, gave us suggestions about what functions they feel necessary for event reporting based on their experience. It is fundamental to address visibility and accessibility in the design which require the elements of the system to be visible and easy for users to determine the potential actions and to evaluate the current status of the system [23]. In the web page design, we adopted a straightforward style with a concise menu and navigation bar to each function and module. A *text prediction* function was added to facilitate users' text entry and searching in the system. For reporting fall cases, a *logic branching* was implemented based on the Common Formats [16] to guarantee the completeness of the reports and prevent irrelevant sub-questions. The thirteen multiple-choice questions make the reporting much easier for reporters to understand what is expected in a report. To collect new pieces of knowledge beyond the pre-defined options, feedback boxes were also provided at the end of each question. In summary, all these features were designed to increase the engagement and satisfaction of reporters.

Relevance determines ultimate training value since even the best training could be a waste of resources if the participants fail to practice in their everyday work [10]. To ensure the supported knowledge is useful and practical, the knowledge base of our system was established on evidence-based solutions. We matched the actionable solutions to according question options in the Common Formats. Once a reporter chooses a certain option, the corresponding solutions to this option will be provided, which makes the provided knowledge support highly relevant to the reported cases. For example, if a reporter chose "No" for the question asks if fall risk assessment was done for the patient, the solutions regarding to how to assess fall risk for patients will be given. Moreover, our system is expected to become more favorable as the actionable solutions are proved by accumulated feedback of reporters.

Level 2: Learning

Learning is defined as to what degree participants acquire the intended knowledge, skills and attitudes based on their participation in the learning event [10]. This stage emphasizes the learning effects. Several factors are raised in this level and will influence the learning effect. First is knowledge, which is the degree to what the users have learned for certain information, as characterized by the phrase, "I know it". Second is skill, which is the degree to which they know how to do something or perform a certain task, as illustrated by the phrase, "I can do it right now" [10]. Moreover, not only the knowledge and skill are important to staff, but also the attitude and confidence for using what they learned. In some organizations, poor performance of staff was thought due to lack of knowledge and skills, and thus repeatedly and costly trainings

were required. In fact, it is highly possible that the poor performance is due to a lack of motivation or other environmental factors in reality.

In order to achieve the goals in the learning level, our system provides authoritative and relevant solutions from the knowledge base to the reporters. The solutions are targeted and practical since they are provided according to the specific reported event as previously mentioned. This mechanism also solves the most common barrier for current event reporting system, which is lack of effective feedback. The ultimate goal of our system was to make the learning become a continuous process as part of the workflow in healthcare facilities. Traditional training, which often requires the staff to be involved away from their job, may not be as effective as it is supposed to be. About 70% of learning happens during the job [13]. Thus, the learning efficiency is estimated to be greatly improved since our system integrated the training into the healthcare providers' daily work.

Moreover, a similarity-searching module is under development and will be embedded into our system. Classic fall cases summarized in AHRQ WebM&M were annotated by PSNet topics [14] through text mining and integrated into the system. Once a new case is reported, it will be automatically annotated by PSNet topics. Then the system will provide the reporter with previous cases that have similar annotations and their associated solutions, which may help reporters learn more about how to deal with similar situations. The similarity-searching module will also facilitate to build a "shared learning" atmosphere where the reporters can learn from each other based on the accumulation of reported cases.

Level 3: Behavior

This level focuses on monitoring and evaluating whether and how much the participants apply what they have learned after they return to work [10]. This stage aims to fill the gap between knowledge and practice. It overlaps with the learning level, since the improvement of attitude, confidence and commitment can directly increase the consistency between what staff has learned and their behaviors at their routine job. Identifying and critical behaviors of staff are addressed in this stage. Critical behaviors are the few and specific actions that will have the significant impact on the desired results if performed consistently during the job [10]. In our research, the critical behaviors can be regarded as the most important actions that can reduce the occurrence of patient safety events.

In order to highlight the critical behaviors, the solutions in our system are not provided randomly. Instead, they have inherent priorities based on several factors. The most important is the contributing factors, once a case is reported, the system will label the contributing factors to the case. The solutions related to similar contributing factors will be prioritized. Moreover, the solutions from authoritative organizations will be arranged on top of the list. We also categorized all solutions into three subtypes: principle solutions, specific actions for staff, and specific actions for patients. In general, the principle solutions are listed first, followed by the specific actions. Thus, the solutions provided to the reporters are supposed to be the most helpful and actionable ones based on the reporting case. This strategy is expected to maximize the transformation from knowledge to real actions and the efficacy of the final actions.

Monitoring is to ensure the right actions are executed. Direct monitoring is hard to be accomplished without human labors, thus, a specialized group with an elected leader should be set up. The responsibilities of the group are inspecting the omissions and mistakes made by the healthcare providers, and building the rewards mechanism to make sure the learning knowledge has been applied. Patient interview is also an effective method since patients have the intuitive sense about

whether they have been protected from patient safety events. On the other hand, indirect monitoring can be done by the system through supervising and analyzing the reporting history and results. Each reporter should have an independent account under which the behavior of reporters can be analyzed based on a long-term reporting history. More discussion is available in the Results level. The aim of monitoring is to improve the performance of users through identifying their mistakes, increasing the reporting quality, and facilitating users apply what they learned during work rather than punishing the users or their colleagues when they make mistakes.

Level 4: Results

Results are defined as to what degree the targeted outcomes occur as a result of learning from events and subsequent reinforcement [13]. Measuring the results simply by counting the number of the reported cases can be an initial step. Ultimately, an overall evaluation is needed to determine the change of outcomes, i.e. patient fall rate, is due to the application of event reporting system. The Targeted Solutions Tool (TST)® [2], developed by Joint Committee Center for Transforming Healthcare contains features in fall events analysis and strategic planning. Through TST, the accumulated fall cases will be analyzed to create a summary, including top contributing factors for fall cases, primary causes, injury levels, fall locations, etc. And different presentation forms like tables, pie chart, and column chart are given to enhance the clarity and explicitness of the summary.

Level 1: Make the system easy to use	Level 2: Increase users' learning effects
<ul style="list-style-type: none"> Fashionable style of page design. Intelligent searching function. Text prediction function. Logical reporting style. Dynamically generated and updated term vocabulary. Solution entries with both contributing factor label and CF label. 	<ul style="list-style-type: none"> Knowledge base built on authoritative agencies' guidelines. Specific solutions based on user's reporting case are provided. WebM&M cases annotated by embedded PSNet Topics. Similarity case searching function embedded.
Level 3: Promote users' behavior change	Level 4: Monitor system's impact
<ul style="list-style-type: none"> Account based reporting for behavior changing management. Display of annotations for each case. Feedback solutions ranking based on the contribution factors. 	<ul style="list-style-type: none"> Dynamically display of statistical information as soon as user complete the reporting. Track for user reporting, analyze fall events' changing. Collection of users' feedback for similar cases, solutions, etc.

Figure 2 – System features and corresponding aims within four levels of the Kirkpatrick model

To integrate event reporting with TST analysis, a similar module has been designed in our reporting system. The module will track and analyze the rates of all solutions offered as feedback to reporters. An obviously declining frequency of a certain solution over time may indicate that this solution is well executed; otherwise, the solution may need special attention or modification. Then, a connection between the rate of events and rates of provided solutions is expected to be established. By such a design, the influence of our event reporting system can be quantitatively measured. The limitation of this method is that the evaluation can be processed only when the tracking records reach a sufficient amount. Nevertheless, the data will increase significantly as more reporters involved since our system will be integrated into the daily workflow of healthcare providers. Moreover, this method can also monitor whether the reporters have applied what they have learned into their daily work. Another important application of this method is processing the RCA based on the tracking records, in which way organizations can reversely map the solutions to the specific problems. For

example, in fall events, if a solution that asks patient to wear anti-slippery footwear always appears, organizations may have to pay more attention to the quality of the floor. The result evaluation should be processed periodically to update the system and make sure the users keep learning the most up-to-date and essential knowledge. Once the overall patient fall situation has been improved, it will in turn increase staff's sense of accomplishment and encourage them to keep reporting and learning.

All the features and their major aims under each level of the Kirkpatrick model have been summarized in Figure 2. Overall, the four levels of Kirkpatrick model set a standard schema to ensure healthcare providers to use the system willingly and correctly, and improve their performance in real clinical situations. Also, the process will keep the system up to date. The interactions between the system and users will form a loop of “reporting, feedback, improving and evaluation”. Both the system and users will keep learning from each other and improving themselves.

Discussion

Patient fall events in healthcare settings have been a major concern for years. Applications such as guidelines, training programs, and solution kits have been developed to tackle down this problem, but the effects of these applications are still far away from the expectations. The main problems reside in the low training quality for healthcare providers and the lack of consistency between guidelines and practices in daily work.

In order to improve this situation, we have designed and been developing an enhanced patient fall event reporting system with timely knowledge support based on the Kirkpatrick model. The Kirkpatrick model is applicable in healthcare areas since all the trainings have common features needing to be addressed. A successful training should fulfill at least two conditions: the training should teach the right contents; and the training should set mechanisms to ensure people correctly apply what they learned in real work. The Kirkpatrick model is well suited in accomplishing these conditions with the help of the four highly integrated stages defined by the model.

Following the principles from all stages of the model, the design of the knowledge-based reporting system would support basic conditions and maximize the training effect through three key aspects: 1) ensuring the contents in the knowledge base are correct and relevant to fall prevention practice; 2) fostering the positive attitudes among staff, making them confident and willing to apply what they have learned; 3) monitoring and analyzing the outcomes such as the behavior changing of the users, to improve the knowledge base.

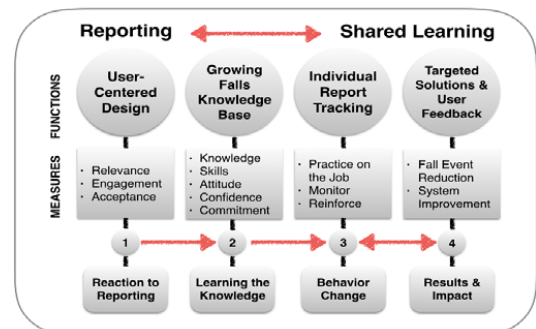


Figure 3 - Design framework of the event reporting system

Our event reporting system was designed for the prevention of patient fall, but not limited to it. The design framework can be generalized to other types of patient safety events in healthcare facilities. The core idea is to provide a user-friendly platform to facilitate clinicians to timely report the patient safety events and integrate the process of learning from errors into their daily workflow. The whole design framework of the event reporting system was abstracted in Figure 3. Overall, reporting is merely an initial step toward shared learning. The four levels are interlocked with each other. The first two levels aim to make the system easy to use, and let users learn what they need. They are the foundations for the remaining two levels, which aim to facilitate and monitor the behavior change of users and measure the final outcomes. The improved outcomes, will in turn, encourage the reporting and learning. A virtuous cycle is expected to be established.

Future work will mainly focus on solving the limitations of our event reporting system at current stage. The limitations include: 1) It is difficult to keep updating the knowledge base without human labors over time. 2) Some advanced functions for data storage and analysis have not been fully built into the system yet. These functions are important for the RCA of reported cases and measurement of the outcomes. 3) Using Common Formats as the template to collect data has advantages for its simplicity and explicitness, but it is not comprehensive enough to cover all the details of fall cases, for example, patients' medication usage. The loss of details will decrease the specificity of the solutions provided by the system. To overcome this limitation, we will improve and standardize the reporting form for patient falls by integrating Common Formats and TST®. 4) The experts of patient safety may have divergent views during the evaluation of our knowledge base. Further evaluation involving diverse expert groups has been planned. 5) The solutions in our knowledge base may not be applicable to all healthcare facilities, since the facilities may have different scales, financial budgets, and types of patients, etc. We plan to set individual accounts for different users to grow a personalized knowledge base. This mechanism could encourage the reporters to interact with the system instead of passively accepting the learning materials, thus both staff and the system are always in the "personalized" learning process. Also, the usability and field tests are necessary, which let target users to use the system to figure out potential problems and evaluate the system's effects on them.

Conclusion

The functions and designed features of our fall event reporting system satisfy the needs of fall management and healthcare staff training. The integration of the Kirkpatrick model makes the system promising to fix the gap between knowledge and practice, thus can better improve the severe fall situations in healthcare organizations. In the future, we will continue improving our system to overcome the current inadequacy and limitations.

Acknowledgements

The study is supported by a research grant (1R01HS022895) from the Agency of Healthcare Research and Quality, UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), and University of Texas System Grants Program (#156374).

References

- [1] D. Oliver, F. Healey, and T.P. Haines, Preventing falls and fall-related injuries in hospitals, *Clin Geriatr Med* **26** (2010), 645-692.
- [2] J. Commission, Preventing falls and fall-related injuries in health care facilities, *Sentinel event alert/Joint Commission on Accreditation of Healthcare Organizations* (2015), 1.
- [3] C.A. Wong, A.J. Recktenwald, M.L. Jones, B.M. Waterman, M.L. Bollini, and W. Dunagan, The cost of serious fall-related injuries at three Midwestern hospitals, *The Joint Commission Journal on Quality and Patient Safety* **37** (2011), 81-87.
- [4] T.P. Haines, A.M. Hill, K.D. Hill, S.G. Brauer, T. Hoffmann, C. Etherton-Beer, and S.M. McPhail, Cost effectiveness of patient education for the prevention of falls in hospital: economic evaluation from a randomized controlled trial, *BMC Med* **11** (2013), 135.
- [5] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, *To err is human: building a safer health system*, National Academies Press, 2000.
- [6] J.C. Pham, T. Girard, and P.J. Pronovost, What to do with healthcare incident reporting systems, *Journal of public health research* **2** (2013), 27.
- [7] M.J. Kingston, S.M. Evans, B.J. Smith, and J.G. Berry, Attitudes of doctors and nurses towards incident reporting: a qualitative analysis, *Medical Journal of Australia* **181** (2004), 36-39.
- [8] I. Mitchell, A. Schuster, K. Smith, P. Pronovost, and A. Wu, Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'To Err is Human', *BMJ quality & safety* **25** (2016), 92-99.
- [9] J.J. Waring, A qualitative study of the intra-hospital variations in incident reporting, *Int J Qual Health Care* **16** (2004), 347-352.
- [10] J.D. Kirkpatrick and W.K. Kirkpatrick, *An Introduction to the New World Kirkpatrick® Model* in, Kirkpatrick Partners, 2015.
- [11] D.L. Kirkpatrick, *Evaluating Training Programs: The Four Levels* Berrett-Koehler Publishers, Inc., 2005.
- [12] A. Smidt, S. Balandin, J. Sigafos, and V.A. Reed, The Kirkpatrick model: A useful tool for evaluating training outcomes, *Journal of Intellectual and Developmental Disability* **34** (2009), 266-274.
- [13] J. Kirkpatrick, The Kirkpatrick Four Levels: A Fresh Look After 50 Years 1959-2009, *Training Magazine* (2009).
- [14] Agency for Health Care Research and Quality WebM&M Cases & Commentaries., in, 2016.
- [15] Center for Patient Safety: <http://www.centerforpatientsafety.org/>, in, 2017.
- [16] Patient Safety Event Report - Hospital: Fall, in, Agency for healthcare Research and Quality, 2016.
- [17] A. Terry, C. Mottram, J. Round, E. Firman, J. Step, and J. Bourne, A safer place for patients: learning to improve patient safety, (2005).
- [18] A. Basu, D. Gopinath, N. Anjum, and S. Hotchkies, Feedback in incident reporting-more needed, *Clinical Governance: An International Journal* **14** (2009), 38-41.
- [19] J. Whitaker and F. Ibrahim, Incident reporting feedback experience in a UK secondary care setting, *The Bulletin of the Royal College of Surgeons of England* **98** (2016), 82-84.
- [20] J.E. Anderson, N. Kodate, R. Walters, and A. Dodds, Can incident reporting improve safety? Healthcare practitioners' views of the effectiveness of incident reporting, *International journal for quality in health care* **25** (2013), 141-150.
- [21] S.M. Evans, J. Berry, B. Smith, A. Esterman, P. Selim, J. O'Shaughnessy, and M. DeWit, Attitudes and barriers to incident reporting: a collaborative hospital study, *Quality and Safety in Health Care* **15** (2006), 39-43.
- [22] K. Martowirono, J.D. Jansma, S.J. van Lwijk, C. Wagner, and A.B. Bijnen, Possible solutions for barriers in incident reporting by residents, *Journal of evaluation in clinical practice* **18** (2012), 76-81.
- [23] C. Abras, D. Maloney-Krichmar, and J. Preece, User-centered design, *Bainbridge, W. Encyclopedia of Human-Computer Interaction*. Thousand Oaks: Sage Publications **37** (2004), 445-456.

Address for correspondence

Yang Gong, MD, PhD

7000 Fannin St. Suite 600, Houston, Texas 77030, USA

Yang.Gong@uth.tmc.edu; 713-500-3547

IV. Knowledge Management

This page intentionally left blank

Managing User Needs During the EHR Implementation in Buenos Aires City: The HelpDesk Role

L. Alassia^a, C. Palermo^a, F. Recondo^a, M. Giussi^a, A. Stieben^a,
A. Baum^a, F. Gonzalez Bernaldo de Quiros^b

^a Dirección de Informática Clínica, Estadística y Epidemiológica, Health Ministry, Buenos Aires City

^b Health Cabinet Adviser, Health Ministry, Buenos Aires City

Abstract

Enablement, guidance, and proactive preparation for a new IT system implementation has proven to be a smart way to prepare people to embrace change. These activities are closely related to change management approaches which seek to help people understand and adopt specific behaviors. That being said, investing in such activities becomes the cornerstone of the success of the project. Leading IT companies around the world include in their strategy for new services and offering deployments, a proactive HelpDesk service. This not only helps build long lasting/trusted relationships between end users and IT sectors but also helps reduce cost and maximizes the Return of Investment. A streamlined process and easy to use/fluent communication channel between parties are powerful risk management/Quality Assurance and Continuous Improvement tools. In this paper, we address the example of a HelpDesk support team implementation of a city-scaled Electronic Health Records implementation.

Keywords

Health Information Systems; Workflow; Argentina

Introduction

HelpDesk support is key to set in motion a successful implementation of any IT system. Not only does it help with User Experience and Satisfaction it also helps with the Return of Investment (ROI) associated with the funds allocated to the IT system. A well-qualified HelpDesk would be not only fix technical issues, but would also ensure root causes are tackled accordingly and prevented from reoccurring in the future. HelpDesk employees perform these tasks working side by side with developers. An IT system implementation is not just a junction of applications. There is consensus among the private and public sectors a successful implementation depends on organizational factors [1,2], including the ability to answer to the needs of users directly or indirectly involved in the implementation process.

An information system attempts to accurately reflect and streamline processes and, also, bring them to life as they are executed. The IT system is meant to prove it's worth by adding value to the process, positively impacting the overall Organization [3-6]. Implementation reports and practices show there is a continuous life cycle adaptive evolution between work processes and the applications of an information system, requiring appropriate change management in the first as well as the establishment of continuous improvement cycles [2, 7]. In this context, Helpdesk implementations - understood as a tool leveraged to support the operational line - is part of the Organizational Strategy. A Helpdesk combines technologic and human resources, providing services to manage and solve

incidents/scenarios and/or requirements related to Information and Communication Technologies (ICT). Clinical Information Systems (SCI) involve several ICTs as well as different types of users, who need adequate support and training to ensure the successful use of such systems [8].

In addition to the HelpDesk's benefits to supporting the operational aspects through a fluid communication with the central-level and by contributing to increase user experience/satisfaction, HelpDesks also act as management tools. It allows management to gather feedback from end users regarding areas of improvement, recurrent scenarios, automation opportunities, etc. empowering management to sort/label problems and allow through reporting, accurate and up-to-date information to be available for leadership's smart decision making, both for corrective and preventive actions [9]. This paper's objective is to describe a HelpDesk implementation process and its benefits to operational management and excellence applied to the primary care level of Buenos Aires' public health system.

Methods

Design

This is a cross-sectional study. A secondary analysis of HelpDesk's database was performed.

Setting

The Autonomous City of Buenos Aires (CABA) is Argentina's capital city. According to the 2010's national census, the population of the city is 2,890,151, and the population of its urban cluster/outskirts, locally named Greater Buenos Aires, is around 12,801,364 making Buenos Aires the largest urban area of the country and the second largest one from South America [10].

Health Care Network and Public Health System in CABA

Three sub-systems coexist in Argentina's Health System. In addition to the Public sub-system, there is health care coverage derived from Social Security, which assists the population with formal jobs and, lastly, private health coverage, funded with contributions from its users. The public health sub-system is funded by the government (i.e., local, provincial or national) and provides health care for free to the population. The CABA health network has 33 Hospitals, 44 primary and community health centers (CeSAC) and neighborhood medical centers which are divided into four regions. Each health region is subdivided into areas which include a general hospital and the local CeSAC. The CeSACs are part of the city's primary health care network (APS), which provides services to approximately 500,000 medical queries per year, 50% of them correspond to

the pediatricians' specialty and the other half are carried out by primary care physicians and OBGYNs. In addition, CeSAC provides services to approximately 260,000 non-medical specialization queries, including nutrition, social work, psychology, obstetrics, and dentistry. Based on data from 2014, 17.8% of the population living within CABA receive assistance only from this system. This percentage rises to 31.2% in areas South of the City. The network also takes care of people residing in the outskirts of Buenos Aires metropolitan area and in other provinces, although to a lesser extent [11].

Implementation Process and Management

Fifty-two days passed from the first to the last link between CeSAC and the HelpDesk (4/21-6/13). The tickets were labeled based on: problems (Category) and priority (low, medium, high, and critical). Given some tickets could not be included within existing categories, HelpDesk managers had ownership of the generation of new categories, making them available for future similar scenarios.

Results

Until November 14th, the final cut-off date for this study, 1400 requests were made to the HelpDesk, coming from the 44 primary care centers of the City of Buenos Aires. The overall analysis of these orders is shown in Table 1, which lists the total number and percentage of tickets sorted by the main categories.

Table 1– Generated Tickets to November 2016

Category	N	% (n=1400)
Supplies	367	26.21%
Patient Indexing	318	22.71%
Electronic Schedules	210	15%
Maintenance	203	14.50%
Informatics Support	87	6.21%
Process Support	68	4.86%
Pharmacy	38	2.71%
Other	31	2.21%
Security	23	1.64%

Considering the context of progressive implementation, we also analyzed the qualitative difference in terms of ticket categories and the rate of ticket generation, adjusted by business days in the month. This analysis shows the fluctuation of the requests, motivating the ticket generation routed to the central level, quickly change. Regarding the labeling of the reported problems, in the first two months of implementation, the root cause of tickets generated was related to legacy issues (e.g., supply orders and maintenance issues), a trend that changed from the third month onwards to issues related to operational support of the actual implementations (Patient Indexing and electronic scheduling represent almost 40% of orders). Figure 1 illustrates the counts of the most used categories in the first 3 months compared to the 7th month.

As for the number of tickets generated per day, the value accounts for a mildly increasing trend during the initial months with a steep increase in September, keeping the increased rate until the cut-off date. This trend is best observed when adjusting

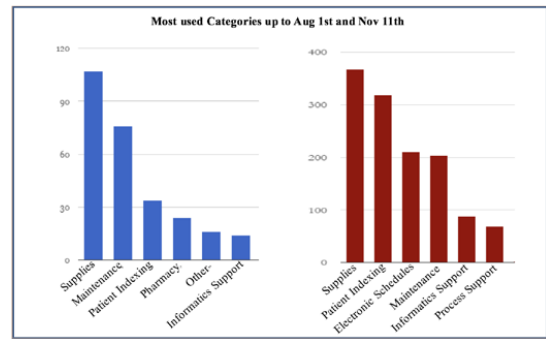


Figure 1 - Most used categories in the first 3 months and at the 7th month.

the number of orders per working days of the month (Table 2), going from an average of 7-8 orders/day in August to 23 orders/day in November. This phenomenon correlates with computerized processes enabling remote support.

Table 2 – Tickets by Month (workdays)

Month	N	Tickets/Workdays
April-May (from 4/21)	58	2
June	142	6.45
July	102	4.86
August	175	7.61
September	324	14.73
October	357	17
November (untill 11/14)	237	23.7

Discussion

As previously mentioned, a help desk has different capabilities and purposes, depending on the kind of users under consideration within scope. Analyzed in the context of the present implementation, the HelpDesk is used as a change management tool. In addition to serving as a sensor of operational needs, the HelpDesk also becomes an indirect tool to satisfy, totally or partially, the most pressing needs of health centers.

From a management point of view, the central level represented by the Ministry could quickly diagnosis the needs in the first level of care. This led to operational decisions aimed at formalizing the processes to resolve these problems, outlining specific responsibilities where applicable, and establishing supervision over external third party vendors (i.e. outsourced). With regards to historical problems, most were reported in the first few months; action was triggered by the central level focusing on their resolution, especially on maintenance issues, which were addressed mostly through regular follow-up meetings with the office accountable for the third parties' supervision. We then proceeded to standardize response times for the most frequently incoming tickets (i.e., supply chain and basic maintenance issues), which allowed the generation of continuous improvement cycles of issue resolution quality assurance processes.

In alignment with the progress of the overall implementation, the HelpDesk also becomes the channel for reporting errors that require corrective actions as well as suggestions for procedural or computer improvements as a result of user empowerment.

The present work has several limitations. Because of being a cross-sectional study, the results should be interpreted as being potentially impacted by factors not being considered within the scope of this study. On the other hand, although all the CeSACs have the HelpDesk implemented, half of them have not implemented HCE applications and electronic scheduling. The mentioned scope corresponds to the 2017 agenda. Finally, the user satisfaction was not measured with a qualitative methodology. This could have enriched the work presented in this paper.

Conclusion

As the Health Informatics Project for Buenos Aires City advances, we conclude that the HelpDesk is a valuable tool, not only for management, but also for change facilitators. Results in our study will be used to plan further implementations as we move forward throughout different settings.

References

- [1] J. Aarts, Understanding implementation: the case of a computerized physician order entry system in a large dutch university medical center, *J Am Med Informatics Assoc.*, **11** (2004): 207-216.
- [2] M. Cucciniello, I. Lapsley, G. Nasi, C. Pagliari, Understanding key factors affecting electronic medical record implementation: a sociotechnical approach, *BMC Health Serv Res* **15** (2015):268.
- [3] T. Kelay, S. Kesavan, R.E. Collins, J. Kyaw-Tun, B. Cox, F. Bello, R.L. Kneebone, N. Sevdalis, Techniques to aid the implementation of novel clinical information systems: a systematic review **11** (2013):783-791.
- [4] E.M. Campbell, K.P. Guappone, D.F. Sittig, R.H. Dykstra, J.S. Ash, Computerized provider order entry adoption: implications for clinical workflow, *J Gen Intern Med* **24** (2009):21-26.
- [5] D. Martin, J. Mariani, M. Rouncefield, Managing integration work in an NHS electronic patient record (EPR) project, *Health Inform J* **13** (2007):47-546.
- [6] M. Berg, C. Langenberg, I. Berg, J. Kwakkernaat, Considerations for sociotechnical design: experiences with an electronic patient record in a clinical context, *Int J Med Inform* **52** (1998):243-251..
- [7] M.J. Van Der Meijden, J. Huibert, J. Tange, J. Troost, A. Hasman, Determinants of success of inpatient clinical information systems: a literature review, *J Am Med Inform Assoc* **10** (2003):235-243.
- [8] D. Capurro, M. Soto, P. Giacaman, S. Catalán, Characterization of help desk issues after the implementation of an emergency department electronic health record *Stud Health Technol Inform* **216** (2014):875-875.
- [9] D. Giunta, L. Pintos, A. Stieben, G. Garcia, A. Briatore, A. Bertioia, S. Benítez, D. Giunta, A. Baum, F. Gonzalez Bernaldo de Quiros, Health information systems: evaluation and performance of a help desk health information systems: evaluation and performance of a help desk *Stud Health Technol Inform*, **210** (2015): 536-540.
- [10] Gobierno de la Ciudad de Buenos Aires. Ciudad de Buenos Aires. 2016 Available at: <http://www.buenosaires.gob.ar/laciudad/ciudad>.
- [11] G.M. Bordoni, F. Plazzotta, A. Baum, I.C., G.F. Bernaldo de Quiros, Elaboración e implementación de una Agenda Digital en Atención Primaria en la Ciudad Autónoma de Buenos Aires, *CBIS* 2016.

Address for correspondence

Leandro Noer Alassia

lalessia@buenosaires.gob.ar

+5491152572102

Expressing Biomedical Ontologies in Natural Language for Expert Evaluation

Muhammad Amith MS^{a,*}, Frank J. Manion MS^{a,b*}, Marcelline R. Harris PhD^{b*},
Yaoyun Zhang PhD^a, Hua Xu PhD^a, Cui Tao PhD^a

^a School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas, United States

^b University of Michigan, Ann Arbor, Michigan, United States,

* Equal contribution

Abstract

We report on a study of our custom Hootation software for the purposes of assessing its ability to produce clear and accurate natural language phrases from axioms embedded in three biomedical ontologies. Using multiple domain experts and three discrete rating scales, we evaluated the tool on clarity of the natural language produced, fidelity of the natural language produced from the ontology to the axiom, and the fidelity of the domain knowledge represented by the axioms. Results show that Hootation provided relatively clear natural language equivalents for a select set of OWL axioms, although the clarity of statements hinges on the accuracy and representation of axioms in the ontology.

Keywords:

Natural Language Processing; Biomedical Ontologies; Knowledge Management

Introduction

Ontologies are artifacts of encoded knowledge that represent pieces of information in a *subject > predicate > object* format (e.g., *HPV virus > increases risk for > cervical cancer*) known as triples. Ontologies aim to represent a defined domain space using interlinked triples, harnessed by machines for further processing or machine intelligence tasks. For ontologies to be machine-readable, special syntax is utilized to encode the interlinked triples. For example, the Web Ontology Language (OWL)¹, or Resource Description Language (RDF)² / Terse RDF Triple Language (Turtle)³ are commonly used to encode the interlinked triples.

As a data-source, ontologies are not immune to errors or inconsistencies. Reasons for the errors and inconsistencies are beyond the discussion of this paper, but the importance of this area of ontology work has been highlighted recently in [1] and [2]. Additionally, most biomedical ontologies on the National Center for Biomedical Ontologies (NCBO) Biportal do not report any evidence of any evaluation [3].

Background

Evaluation frameworks are invaluable to knowledge engineers constructing or refining an ontology, and when assessing if a given ontology is fit for use. According to Gómez-Pérez, ontology evaluation falls under two categories: validation and verification [4]. Validation examines the purposeful, extrinsic aspect of the ontology while verification examines the internal aspects (e.g., the quality of terms, graphical structure, etc.).

Evaluation of ontologies typically involves assessment along three axes: Syntax, Semantics, Pragmatics. Both qualitative [5-7] and quantitative [8] assessment of the three axes are possible [9]; a common approach is to enlist subject matter experts to review the ontology artifact. Our focus is on the veracity of ontologies, which we ascribe as a verification-based evaluation. We presume that Subject Matter Experts (SMEs) are integral to the quality of the ontology during development phases. However, for an SME to review and assess an ontology, specifically those with little to no knowledge engineering background, we need to present it in a format that is accessible.

Among the challenges when engaging subject matter experts is the substantial learning curve to help these experts interpret the knowledge representation in the ontology [10; 11], and their lack of familiarity with ontology construction and visualization tools such as those in the commonly used ontology editor Protégé. This is unfortunate as many experts could provide significant input in improving the quality of the ontology. One possible solution is to translate the knowledge in the ontology to human-readable natural language statements. Below, we discuss details of our approach.

We propose that expressing an ontology in natural language is likely to provide a more readily understandable approach to interpret the interlinked triples, and thereby provide a valuable resource when engaging domain experts in working with the ontology. The natural language sentences produced by the Hootation tool (described below) can be used to assess ontologies along all three axes. Syntax can be assessed by determining if statements constructed from the model and expressed in natural language are correct when compared to uses cases and other artifacts of importance within the domain. Semantics can subsequently be assessed by determining if the definitions of labels expressed in the sentences convey an accurate and complete meaning in the context of their intended use, and if the classes, association, attributes and relationships in the sentences are understandable and relevant. Finally, close examination of the sentences produced can reveal pragmatic issues with the ontology such as formal completeness i.e., what may be missing from the ontology, and what cognitive effort on domain experts is needed in understanding the ontology. The latter is sometimes used as a proxy measure for consistency of the ontology.

“Hootation” Java Library

Related Studies

Natural Language Generation (NLG) is an expansive topic that has been the focus of considerable previous research. NLG is one of the two main topics of natural language processing – the other centers on natural language understanding (NLU), which is the focus of much biomedical

¹ <http://www.w3.org/TR/owl2-overview/>

² <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>

³ <https://www.w3.org/TeamSubmission/turtle/>

informatics research. While NLU centers on interpreting free text into data for machines to understand, NLG focuses on interpreting data from the machines into free text or documents for humans to understand. In the context of this study, the emphasis is on transforming triples from ontological models to natural language statements that would help evaluate the knowledge contained in biomedical ontologies. The merits of NLG applications for biomedical ontologies include question-answering, document creation and summarization from datasets, concealing the complexity of the syntax, and ontology evaluation.

An early work in this area was ModelExplainer [12] that generated lines of text from object oriented models. Other relevant work for authoring and NLG applications involved the use of bi-directional Controlled Natural Language (CNL) for OWL 1.1 such as Attempto Controlled English (ACE) [13], Sydney OWL Syntax (SOS)[14], Rabbit [15] etc. However, CNL are compounded by the issues of NLU (ambiguity of text) and NLG (difficult to comprehend for users and limitations the label's nomenclature). None of what has been described provides a dedicated OWL2-to-NL engine that is portable for application use. However, NaturalOWL employs a basic template approach, but it depends on a separate authoring tool for domain dependent generation [16; 17].

Hootation

Our Hootation software library is derived from the natural language generation work by Agile Knowledge Engineering and Semantic Web (AKSW) Research Group [18; 19] developed initially for a semantic web application. The NLG layer harnesses the OWL-API; as initially developed, it supported the translation of 12 logical axioms for OWL2. We added support for 6 additional logical axioms, with plans to add more translation for the remaining axioms. Hootation also utilizes SimpleNLG [20], a state of the art NLG engine that provides flexible APIs to manipulate morphological and syntactical aspects of a generated statement. SimpleNLG also allows the use of the NIH Specialist Lexicon [21] for expanded coverage for medical lexicon, which we have yet to exploit, but an added benefit for biomedical ontologies.

While many NLG applications focus on producing documents or other large bodies of text, an immediate goal is proper translation of each individual axiom to NL statements, so that biomedical experts can rate the veracity of the information and then report on the content quality of the ontology. We intend to integrate the Hootation API library into our continued work to provide a web-based tool for comprehensive ontology evaluation (see Future Direction section for details). Source code and a Java binary library will be available for open source distribution⁴.

Methods

Our primary objective was to determine whether Hootation could accurately produce natural language from biomedical ontologies in a way that is understandable for subject matter experts. We also wanted to evaluate factors that contribute to or hinder the clarity of the natural language. Java code was developed to interface with the Hootation API library, and output for each of the ontologies was exported in CSV format including, for each natural language statement, the corresponding axiom in OWL Manchester format and the type of logical axiom.

Sample

Because most ontologies do not use every axiom type available by OWL, three ontologies were used to capture the NL translation for diverse axioms.

The "People" Ontology represents knowledge on the types of people based mostly on familial information. The People ontology is a teaching tool for University of Texas Health Science Center students, used as an introduction to the development OWL-based ontologies and as an introduction to the descriptive logic power of OWL. This ontology is based on descriptive definitions from California Polytechnic State University, in San Luis Obispo, California [22]. The ontology used for this study contained 13 classes, 8 properties, 9 instances (90 total) and a variety of axiom types. This ontology was included because of the simple and universal nature of the encoded information as well as its utilization of various axiom types.

The "Informed Consent Ontology" (ICO) [23] is a preliminary ontology based on the analysis of informed consent templates and blank informed consent forms obtained from two separate Institutional Review Boards (IRB) at the University of Michigan. In its current iteration, ICO focuses on informed consent documents and processes. Consequently, the concepts represented by classes and relations in the work are in the context of informed consent documents, and recommendations for addressing concepts of risk, privacy, and other notions of precepts laid out in US Common Law and medical ethics. ICO is based on the Basic Formal Ontology (BFO) [24], represented in the Web Ontology Language (OWL2), was built on Open Biomedical Ontologies (OBO) Foundry principles [25], and inherits the classes, relations, and axioms from the Ontology of Biomedical Investigations (OBI)[26]. ICO contains 375 classes, including 163 ICO-specific classes and 86 properties. The ontology contains 677 axioms, however many of these axioms are inherited directly from the OBI framework, leaving 183 ICO-specific axioms that are studied in this paper.

The Time Event Ontology (TEO) is a derivative of the Clinical Narrative Temporal Relation Ontology (CNTRO) by the Ontology Research Group at the School of Biomedical Informatics (University of Texas Health Science Center) [27]. TEO contains entities and definitions relating to temporal information and their semantic relationships between them. Its intention is to "provide a formal conceptualization of temporal structures in both structured data and textual narratives" and "core semantic components for representing temporal events and relations to enable reasoning capacities in temporal relations." TEO (version 1.7) contains 156 classes, 51 properties, 8 instances, and 1026 axioms. Similar to ICO, TEO is based on BFO.

Evaluation Procedures

For each ontology, two persons familiar with the logic of the ontology evaluated the NL expression along three dimensions. First, *Clarity* was scored from 1-3, where 1 indicates the natural language expression of the axiom is clear, unlikely to cause confusion or ambiguity, 2 indicates the natural language expression of the axiom is clear, but there may be ambiguity attributable to the axiom, and 3 indicates the natural language expression is not interpretable. *NL Fidelity to Axiom*, addressed whether the natural language expression demonstrated fidelity to the underlying axiom (i.e. logic was accurately expressed in the natural expression). If this dimension was scored as "yes", the natural language expression is an accurate reflection of axiom; if "no", the tool appears to have misinterpreted the logic. *Axiom Fidelity to Domain*, addressed the fidelity of the axiom itself to domain

⁴ <https://goo.gl/sGkYli>

knowledge. Although an evaluation of the fidelity of the axiom to domain knowledge is an assessment of the ontology and not an evaluation of the expression generated by the Hootation tool, it addresses this dimension because both syntactic and semantic issues within the ontology itself sometimes confounded assessments of the clarity of the NL expression. A score of 1 indicated that the reviewer agrees with axiom, 2 that the reviewer disagrees or is uncertain about concepts *or* relationships in axiom, and 3 that the reviewer disagrees or is uncertain about concepts *and* relationships indicated by the axiom. Overall, familiarity with the ontology is important as it yields direct expertise of the intention and construction of the axioms, and adeptness in assessing the translation based on the three dimensions.

Evaluators used online, shareable spreadsheets to record their assessments of each dimension. The first three columns of the spreadsheet presented the axiom type (e.g., SubClassOf, EquivalentClasses), the axiom logic expressed in description logic notation, and the natural language expression generated by the Hootation tool (See Table 1). After reviewing the content of each row, each of the evaluators assigned to the specific ontology recorded their assessment of each of the three dimensions in separate columns. Reviewers were not blinded to each other's assessments because the goal was informed critique and convergence on evaluations of each natural language expression. Disagreements were recorded if scoring decisions could not be reconciled between reviewers.

Table 1 - Sample output

Axiom Type	Logical Axiom	NL Equivalent
SubClassOf	ICO_0000062 \sqsubseteq ICO_0000073	every human subject unable to give informed consent is a human subject

In order to explore relationship when the NL statement may be clear but the axiom is incorrect, we utilized IBM SPSS (v23) to calculate Spearman's rho correlation between the *Clarity* values and *Axiom Fidelity to Domain* knowledge values.

Results

Interrater agreement was calculated among the raters, and the overall agreement for the aforementioned metrics were 86% for Clarity, 91% for Fidelity to Natural Language, and 90% for Fidelity to Domain. For the People ontology, the agreement was 83% for Clarity and Fidelity of Natural Language to Axioms, and 98% for Axiom Fidelity to Domain. Likewise, for Time Event and Informed Consent the agreements were 82%, 96%, 88%; and 92%, 93%, 85%, respectively. Overall, there was high agreement with the results of the assessment. We caution that each ontology was independently assessed and consequently, the results do not yield normalized quality data of the underlying ontologies.

The data in Table 2 and Table 3 were aggregated and segmented to comprehensively evaluate the evaluators assessment of Hootation's results.

Table 2 provides a summary of the overall averages for each of the assessment described earlier. Across the three ontologies, the People ontology demonstrated the best rating for Clarity and Fidelity to Domain knowledge; the Clarity metric rating was 1.19, and Axiom Fidelity to Domain was 1.01. The ICO ontology demonstrated the best NL Fidelity for the axioms (95%). For the Time Event ontology, the average assessment scores were 1.32 for Clarity, 92% for NL Fidelity to Axiom, and 1.13 Axiom Fidelity to Domain. The Informed Consent ontology average assessment scores were 1.28 for

Clarity, 95% for NL Fidelity to Axiom, and 1.36 for Axiom Fidelity to Domain). Altogether, the three ontologies yielded an average Clarity of 1.26, 92% (yes) for NL Fidelity to Axiom, and 1.17 for Axiom Fidelity to Domain. The correlation between clarity and fidelity to domain knowledge reveals a positive, strong linear relationship that was statistically significant ($r = 0.76$, $p < 0.01$).

Table 2 - Average assessment of ontologies

Ontology	Clarity (μ , σ)	NL Fidelity to Axiom (% Yes)	Axiom Fidelity to Domain (μ , σ)
People	1.19 (0.42)	90 %	1.01 (0.14)
Time Event	1.32 (0.63)	92 %	1.13 (0.38)
Informed Consent	1.28 (0.58)	95 %	1.36 (0.64)
Average	1.26	92 %	1.17

Table 3 presents data segmented by 14 OWL axioms supported by Hootation. Because of space constraints, below we show only the two examples demonstrating the range of metrics observed related to axioms. See <https://goo.gl/Br3qQU>

for full table. The majority of the axioms types were SubClassOf, and others appearing in two or more ontologies were ClassAssertion, EquivalentClasses, ObjectPropertyDomain, ObjectPropertyRanges, DataPropertyDomain, DataPropertyRange, DisjointClasses, and SymmetricalObjectProperty. Five axiom types, FunctionalDataProperty, FunctionalObjectProperty, DataPropertyAssertion, DifferentIndividuals, and ObjectPropertyAssertion, only appeared in the People Ontology. Not all ontologies used every axiom types, and variations in Clarity, NL Fidelity to Axiom, and Axiom Fidelity to Domain are noted across axiom types. For example, across all three ontologies, more complex axiom types such as ObjectPropertyRange scored much worse on clarity than less complex axioms such as SubClassOf.

Table 3 - Assessments by Axiom Type (example)

	Axiom Type	
	SubClassOf	ObjectPropertyDomain
NL Expressions (n)	306	10
People	11	4
Time Event	118	5
Informed Consent	117	1
Clarity (μ , σ)	1.14, 0.52	1.3, 0.73
People	1.00, 0	1.00, 0
Time Event	1.19, 0.49	1.00, 0.63
Informed Consent	1.23, 0.55	3.00, 0
Fidelity of NL to Axiom (% Yes)	99	63
People	100	100
Time Event	100	90
Informed Consent	97	0
Fidelity of Axiom to Do- main (μ , σ)	1.13, 0.58	1.3, 0.73
People	1.00, 0	1.0, 0
Time Event	1.06, 0.29	1.2, 0.63
Informed Consent	1.34, 0.69	3.0

Discussion

The results of this study are encouraging with respect to NLG generation for use in ontology evaluation. Overall, Hootation appears to generate natural language statements with clarity, and fidelity to the axiom.

Problems in clarity included the introduction of mid-level noun phrases that would not be typical in a purely natural discourse between domain experts. For example, the ICO axiom: $\text{ICO_0000171} \sqsubseteq \exists \text{IAO_0000136}.\text{ICO_0000064}$ produced the text *“every answer option text entity is something that is about a study requiring informed consent”*. While technically correct, a more natural English discourse might have been written as *“every answer option text entity is about a study requiring informed consent”*. In general, the introduction of the phrase *“is something that”* in descriptive logic axioms containing an existential restriction of the form $\exists R.C$ caused considerable discussion among the reviewers as to the clarity of the produced phrases.

Another factor we noted during our evaluation that clearly impacts the usefulness of the produced natural language are the class and object property labels of the source ontology. The choice of appropriate labels agreed on by domain experts during the construction of an ontology is generally considered good practice. Iterative use of Hootation during ontology construction and refinement can assist with this.

While not the primary focus of this paper, the fidelity of the axiom to underlying domain knowledge was noted by reviewers of ICO as an issue. Reviewers noted approximately 41 axioms (the number varied slightly between the two reviewers) were not accurate representations of the underlying domain knowledge. ICO developers intend to target these axioms for review and correction. This is strong evidence that tools such as Hootation are useful and effective at improving the quality of ontologies.

Also, the majority of the axiom NL translations were of the SubClassOf type, which intuitively, should be “easy” to translate. However, due to nomenclature of the labels, the translation was not straightforward, and the results point to some lack of clarity and fidelity to domain knowledge.

The finding of a positive correlation between clarity and fidelity of the axiom to domain knowledge merits further investigation, and suggests that when the axiom fidelity to domain knowledge is less accurate, the clarity of the NL statement also diminishes. For example, the TEO axiom, $\text{TEO_0000048} \sqsubseteq \text{TEO_0000084}$, which produced *“every Saturday is a week day”* is technically correct within the context of TEO axioms (Saturday is modeled as subclass of weekday entity), but as a generated statement it could be misleading to human evaluators because Saturday is typically discussed as a weekend day, not a weekday. Future work on this relationship needs to account for the complexity of the axiom as well as the fidelity of the axiom to domain knowledge. For example, an axiom of the general form $A \sqsubseteq \exists B (\exists C.D)$ is more complex than an axiom of the form $A \sqsubseteq B$. We also intend to add an option to the program to address discourse type as typified by the *“is something that”* issue discussed above, allowing production of text more suited to a domain expert.

Software

Other factors that influenced evaluator ratings were software bugs and the stemming algorithm. The tool does not yet support import of external ontologies, so some NLG statements included unresolved names of entities from external ontologies that were not merged into the ontology file, e.g., *“every duration measurement is an iao 0000032”*, where the iao

0000032 is associated to an entity from the external Information Artifact Ontology. Future work may require that the API automatically downloads and resolves references to external ontologies at runtime. Another problem we noticed was introduced by our utilization of the Porter-Stemmer algorithm, which sometimes unnecessarily decomposed a word to a form that is unrecognizable (Ex. *“data”* became *“da”*). WordNet API software packages offer lemmatization that might be an alternative.

Limitations and Future Direction

The Hootation software library was limited to 14 OWL axioms that can be translated to natural language statements, most of which were carried over from the work of previous developers. In the future, we plan on supporting the translation of the full set of axiom types to provide comprehensive translation of OWL axioms to natural language statements.

Also, we did not separately examine the impact on clarity of BFO-based axioms for those ontologies using BFO as upper level ontology. BFO uses a specific realism-based model to provide a framework for building other ontologies against. Evaluation of ontologies based on the BFO framework demands some familiarity of the underlying BFO model and the precise terms used by BFO for representing terms, classes, and relationships. Finally, we recognize the need to consider complexity of axiom types in future studies.

One of the ongoing projects we are engaged in is to develop a web-based tool (*“OntoKeeper”*⁵)[3] to evaluate published ontologies according to various metrics influenced by [8]. One of the metrics include evaluating the veracity of the ontology from subject matter expert review through an online user interface. To do this, we need to translate the logical axioms to natural language statements to be more “readable” for the experts with little knowledge on how to navigate through an ontology or no knowledge of complexities of the OWL/RDF syntax. The finalization of this work will be to integrate the NL translation component to the web-based tool and publish the API library for future research on biomedical ontology evaluation research.

An impending study will perform an extensive evaluation of specific biomedical ontologies with accomplished biomedical ontology experts to review the NL axioms. That study will also include details on implementation of Hootation and near complete support for majority of the axiom types.

Conclusion

To address the specific need of generating human-friendly interpretation of ontology axioms in natural language, we introduce Hootation. This software library utilizes a combination of the OWL-API and SimpleNLG, as architected by AKSW, to produce basic natural language statements for 14 axiom types. By translating the axioms into natural language statements, we can enlist the participation of domain subject matter experts who can therefore easily review an ontology without the barrier of learning the complexity of knowledge engineering. In the future, we plan on incorporating this library into our prototype software tool OntoKeeper to add a subject matter expert review component. Overall, this work has potential implications for bridging the gap between the expertise of domain experts and encoded knowledge in machine encoded syntax.

⁵ previous called “SEMS” (Semiotic Evaluation Management System)

Acknowledgements

Research was supported by National Institutes of Health under Award Numbers U01 HG009454, R01 LM011829, and the Cancer Prevention Research Institute of Texas (CPRIT) Training Grant #RP160015. Ratings were assisted by Mochine Makdour, Jingcheng Du, and Hsing-Yi Song.

References

- [1] M. Clarkson, Representation of parts within the Foundational Model of Anatomy ontology, in: *Proceedings of the Joint International Conference on Biological Ontology and BioCreative (2016)*, ICBO and BioCreative, Corvallis, Oregon, USA 2016.
- [2] H. Yumak, L. Chen, M. Halper, L. Zheng, Y. Perl, and G. Elhanan, A Quality-Assurance Study of ChEBI, in: *Proceedings of the Joint International Conference on Biological Ontology and BioCreative (2016)*, ICBO and BioCreative, 2016.
- [3] M. Amith and C. Tao, A Web Application Towards Semiotic-based Evaluation of Biomedical Ontologies, in: *International Workshop on Biomedical Data Mining, Modeling, and Semantic Integration: A Promising Approach to Solving Unmet Medical Needs, BDM21 2015, 14th International Semantic Web Conference*, Bethlehem, United States, 2015.
- [4] A. Gómez-Pérez, Ontology evaluation, in: *Handbook on ontologies*, Springer, 2004, pp. 251-273.
- [5] A. Baker, "Simplicity", the Stanford Encyclopedia of Philosophy, (2013).
- [6] A. Gomez-Perez, M. Fernández-López, and O. Corcho, *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, Springer Science & Business Media, 2006.
- [7] J. Yu, J.A. Thom, and A. Tam, Requirements-oriented methodology for evaluating ontologies, *Information Systems* **34** (2009), 766-791.
- [8] A. Burton-Jones, V.C. Storey, V. Sugumaran, and P. Ahluwalia, A semiotic metrics suite for assessing the quality of ontologies, *Data & Knowledge Engineering* **55** (2005), 84-102.
- [9] J. Brank, M. Grobelnik, and D. Mladenic, A survey of ontology evaluation techniques, in: *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*, 2005, pp. 166-170.
- [10] A. Rector, N. Drummond, M. Horridge, J. Rogers, H. Knublauch, R. Stevens, H. Wang, and C. Wroe, OWL Pizzas: Practical Experience of Teaching OWL-DL: Common Errors & Common Patterns, in: *Engineering Knowledge in the Age of the Semantic Web: 14th International Conference, EKAW 2004, Whittlebury Hall, UK, October 5-8, 2004. Proceedings*, E. Motta, N.R. Shadbolt, A. Stutt, and N. Gibbins, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 63-81.
- [11] A. Funk, V. Tablan, K. Bontcheva, H. Cunningham, B. Davis, and S. Handschuh, *Clone: Controlled language for ontology editing*, Springer, 2007.
- [12] B. Lavoie, O. Rambow, and E. Reiter, The modelexplainer, in: *Proceedings of the 8th international workshop on natural language generation*, 1996, pp. 9-12.
- [13] K. Kaljurand and N.E. Fuchs, Verbalizing OWL in Attempto Controlled English, in: *OWL: Experiences and Directions (OWLED)*, Innsbruck, Austria, 2007.
- [14] A. Cregan, R. Schwitter, and T. Meyer, Sydney OWL Syntax-towards a Controlled Natural Language Syntax for OWL 1.1, in: *OWL: Experiences and Directions (OWLED)*, Innsbruck, Austria, 2007.
- [15] G. Hart, M. Johnson, and C. Dolbear, Rabbit: Developing a control natural language for authoring ontologies, in: *European Semantic Web Conference*, Springer, 2008, pp. 348-360.
- [16] I. Androutsopoulos, G. Lampouras, and D. Galanis, Generating natural language descriptions from OWL ontologies: the NaturalOWL system, *Journal of Artificial Intelligence Research* **48** (2013), 671-715.
- [17] D. Galanis and I. Androutsopoulos, Generating multilingual descriptions from linguistically annotated OWL ontologies: the NaturalOWL system, in: Association for Computational Linguistics, 2007, pp. 143-146.
- [18] L. Bühmann, R. Usbeck, and A.-C.N. Ngomo, ASSESS—Automatic Self-Assessment Using Linked Data, in: *The Semantic Web-ISWC 2015*, Springer, 2015, pp. 76-89.
- [19] GitHub - AKSW/SemWeb2NL: Semantic Web related concepts converted to Natural language, in: *GitHub*, <https://github.com/AKSW/SemWeb2NL>, 2015.
- [20] A. Gatt and E. Reiter, SimpleNLG: A realisation engine for practical applications, in: Association for Computational Linguistics, 2009, pp. 90-93.
- [21] J.i. Kazama, T. Makino, Y. Ohta, and J.i. Tsujii, Tuning support vector machines for biomedical named entity recognition, in: *Proceedings of the ACL-02 workshop on Natural language processing in the biomedical domain-Volume 3*, Association for Computational Linguistics, 2002, pp. 1-8.
- [22] Introduction to Ontologies with Protege, in: *Department of Computer Science and Software Engineering - Cal Poly, San Luis Obispo*, <https://wiki.csc.calpoly.edu/OntologyTutorial/wiki/IntroductionToOntologiesWithProtege>, 2011.
- [23] Y. Lin, M.R. Harris, F.J. Manion, E. Eisenhauer, B. Zhao, W. Shi, A. Karnovsky, and Y. He, Development of a BFO-based Informed Consent Ontology (ICO), *Bioinformatics* (2014).
- [24] R. Arp, B. Smith, and A.D. Spear, *Building ontologies with basic formal ontology*, Mit Press, 2015.
- [25] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, and C.J. Mungall, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature biotechnology* **25** (2007), 1251-1255.
- [26] R.R. Brinkman, M. Courtot, D. Derom, J.M. Fostel, Y. He, P. Lord, J. Malone, H. Parkinson, B. Peters, and P. Rocca-Serra, Modeling biomedical experimental processes with OBI, *Journal of biomedical semantics* **1** (2010), 1.
- [27] B.L. Cook, A.M. Progovac, P. Chen, B. Mullin, S. Hou, and E. Baca-Garcia, Novel use of natural language processing (NLP) to predict suicidal ideation and psychiatric symptoms in a text-based mental health intervention in Madrid, *Computational and mathematical methods in medicine* **2016** (2016).

Address for correspondence

For correspondence, please contact, cui.tao@uth.tmc.edu

Eliciting the Intension of Drug Value Sets – Principles and Quality Assurance Applications

Nathan J. Bahr^a, Scott D. Nelson^b, Rainer Winnenburger^c, Olivier Bodenreider^d

^aDepartment of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA,

^bDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

^cStanford Center for Biomedical Informatics Research, Stanford, CA, USA

^dNational Library of Medicine, Bethesda, MD, USA

Abstract

Value sets (VSs) used in electronic clinical quality measures are lists of codes from standard terminologies (“extensional” VSs), whose purpose (“intension”) is not always explicitly stated. We elicited the intension for the 09/01/2014 release of extensional medication value sets by comparison to drug classes from the October 2014 release of RxClass. Value sets matched drug classes if they shared common ingredients, as evidenced by Jaccard similarity score. We elicited the intension of 80 extensional value sets. The average Jaccard similarity was 0.65 for single classes and 0.80 for combination classes, with 34% (27/80) of the value sets having high similarity scores. Manual review by a pharmacist indicated 51% (41/80) of the drug classes selected as the best mapping for a value set matched the intension reflected in that value set name. This approach has the potential for facilitating the development and maintenance of medication value sets.

Keywords:

Vocabulary, Controlled; RxNorm; Algorithms

Introduction

Clinical quality measures and value sets

Healthcare professionals and hospitals must report clinical quality measures (CQM) in order to qualify for additional payments under the Electronic Health Records (EHR) Incentive Programs [1]. CQMs define proportion, ratio, or continuous variable measures that help track the quality of services provided within the healthcare system [2]. To calculate CQMs, healthcare organizations must use codes from standard terminologies in EHR data to identify target populations that meet a given set of criteria. For example, the *beta-blocker therapy for left ventricular systolic dysfunction (LVSD)* measure is defined as the proportion of patients receiving beta-blockers among patients diagnosed with LVSD.

Value sets are lists of codes from standard terminologies (including RxNorm for drugs) that are used to identify concepts such as *beta-blockers* and *LVSD*, which can be used to select appropriate patient cohorts. The National Library of Medicine's (NLM) Value Set Authority Center (VSAC) is responsible for the validation and delivery of value sets.

Extensional vs. intensional value set definition

In practice, most value sets are defined by their extension (i.e., the list of their codes), rather than by their intension (i.e., the properties common to all codes in the value set). The intension of the value set is often only reflected in its name. However, the value set name itself is insufficient to unambiguously define the list of its members. For example, the value set *Beta Blocker Therapy Ingredient* is used in a cardio-vascular CQM,

and beta-blockers used as anti-glaucoma agents would be irrelevant in this value set.

Intensional definitions are critical for the maintenance of value sets. For example, when a new beta-blocker becomes available on the market, it should be added to the beta-blockers value set. If the value set is defined in reference to the drug class (i.e., intensional definition), the new drug will be added to the value set automatically. In contrast, if the list of beta-blockers is established by a pharmacist (i.e., extensional definition), it will need to be periodically revisited to reflect new drugs.

For most value sets currently in use to support CQMs, value set developers have created sets of codes from standard terminologies (e.g., RxNorm for drugs), without explicit documentation of why specific codes were selected. These extensional value sets pose a challenge for validation and maintenance.

Objectives

The objective of this work is to elicit the intension of medication value sets in reference to drug classes. More specifically, we have observed that the value set names generally correspond to: 1) a single drug class, like *Statin*; 2) a drug class in the context of a given disease, like *Antibiotic Medications for Pharyngitis*, and 3) multiple drug classes, like *Ace Inhibitor or ARB Ingredient*. Based on this observation, we propose to use single drug classes, as well as combinations (i.e., intersections and unions) of drug classes to specify an explicit structured intension for medication value sets. A secondary objective is to evaluate the quality of medication value sets by comparison to drug classes. We investigated the 09/01/2014 release of medication value sets for eligible practitioners and hospitals, from the NLM Value Set Authority Center (VSAC), located at <https://vsac.nlm.nih.gov/>. These value sets were created against the October 2014 release of RxNorm.

Related work

Most work on clinical quality measures (CQMs) has focused on their potential for electronically tracking and improving delivery of care [3], issues in validating results from CQMs [4], and accuracy and completeness issues of some CQMs [5; 6]. In this investigation, we specifically focus on the quality of value sets used in CQMs.

In previous work, Winnenburger and Bodenreider [7] assessed the quality of disease value sets by comparing them to the disease classes in the source from which they were derived (e.g., SNOMED CT or ICD10-CM). They hypothesized that concepts in the value set were rooted in one or more ancestor concepts and that these ancestor concepts represent the intension. The extension for a reference value set could then be constructed as the root concepts along with their descendant concepts. For example, the value set *left ventricular systolic dysfunction* would ideally include the concept *left ventricular*

systolic dysfunction and its descendants in SNOMED CT. These techniques provided a framework for evaluating the quality of a value set from a known, structured intension.

We take a similar approach for medication value sets. Of note, while disease value sets can be analyzed in reference to the disease hierarchy provided by SNOMED CT or ICD10-CM, medication value sets require drug classification systems external to RxNorm, namely ATC, MeSH, NDF-RT and DailyMed (described below). Moreover, drug classification systems organize drugs according to various dimensions (e.g., indications, mechanism of action) and multiple combined criteria may be needed to fully characterize a medication value set.

RxNorm and drug classification systems

RxNorm [8] is a standardized nomenclature for medications produced and maintained by the NLM. It provides drug concepts and relations among them. RxNorm concepts are also linked to various drug classification systems through RxNorm's companion resource, RxClass. In this investigation, we leveraged the RxNorm and RxClass application programming interfaces (APIs), available at <https://rxnav.nlm.nih.gov/>. More specifically, we used the RxNorm API, to map various kinds of drug entities to ingredients (e.g., the brand name *Lipitor* to *Atorvastatin*), as drug classification systems generally reference ingredients. We used the RxClass API to associate ingredients with drug classes.

The following drug classification systems were used as a source of drug classes. The *Anatomical Therapeutic Chemical (ATC)* [9] classification system is maintained by the World Health Organization (WHO) for pharmaco-epidemiology purposes. Each ingredient is associated with one or more ATC class. For example, *Atorvastatin* is a member of the class *HMG CoA reductase inhibitors*.

The *Medical Subject Headings (MeSH)* [10] is a controlled vocabulary produced and maintained by the NLM for the indexing and retrieval of the biomedical literature. Its drug descriptors are linked to *Pharmacologic Action (PA)* descriptors which describe mechanisms of action and therapeutic uses. For example *Atorvastatin* has the following pharmacologic actions: *anticholesteremic agents* and *hydromethylglutaryl-CoA reductase inhibitors*.

The *National Drug File Reference Terminology (NDF-RT)* [11] is developed by the Department of Veterans Affairs (VA) Veterans Health Administration and associates ingredients with different pharmacological classes, including chemical structure and diseases for which the drug is indicated. For example, *Atorvastatin* is a member of the disease class *Hypercholesterolemia* (among others).

Finally, *DailyMed* [12] associates ingredients with different pharmacological classes, including the Food Drug Administration's Established Pharmacological Classes (EPC), mechanism of action (MoA), and physiologic effect (PE). Although these associations are also defined in NDF-RT, we used DailyMed because it represents a more authoritative source. For example, *Atorvastatin* is a member of the EPC class *HMG-CoA Reductase Inhibitor*, and the MoA class *Hydroxymethylglutaryl-CoA Reductase Inhibitors*.

Methods

Our approach to eliciting the intension of medication value sets can be summarized as follows (Figure 1). We establish sets of ingredients from drug value sets and from drug classes, and we compare lists of ingredients between value sets and

drug classes. Finally, we perform a quantitative and qualitative evaluation of the elicited intensions.

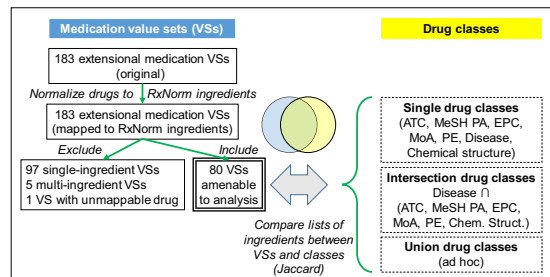


Figure 1 - Overview of the methods.

Establishing sets of ingredients from drug value sets

Medication value sets from VSAC can contain various kinds of RxNorm drug entities, including ingredients (e.g., *Carvedilol*) and clinical drugs (e.g., *Carvedilol 25 MG Oral Tablet*), as well as brand names (e.g., *Coreg*), specific salts and esters (e.g., *carvedilol phosphate*) and other kinds of drug entities. We leveraged the RxNorm API to map the various kinds of drug entities to their corresponding ingredient to simplify the analysis. We excluded multi-ingredient drugs, because the corresponding single-ingredient drugs tend to be listed in the value sets and also because multi-ingredient drugs are not represented consistently across drug classification systems.

Establishing sets of ingredients from drug classes

We leveraged the RxClass API to find the list of RxNorm drug members for drug classes from ATC, MeSH, NDF-RT, and DailyMed. As was done for drug entities from the value sets, we ignored multi-ingredient drugs and mapped the various kinds of drug entities to their corresponding ingredient.

In addition to the (single) classes found in drug classification systems, we created combination classes to represent the sets of drugs reflected in value set names. Namely, we created two types of combination classes: 1) intersection classes, where each single class is intersected with each of the disease classes (attempting to approximate value sets, such as *Antibiotic Medications for Pharyngitis*); 2) union classes, where multiple single classes are merged (attempting to approximate value sets, such as *Ace Inhibitor or ARB Ingredient*). We created the intersection classes systematically for each single class. In contrast, we created union classes corresponding to the best match for each value set (by finding the single class that is most similar to a given value set, and then the best single class that is most similar to the drugs not covered at earlier steps).

Comparing ingredients between value sets and drug classes

We used the Jaccard coefficient to measure the similarity between value sets and (single or combination) drug classes based on their normalized ingredients. The Jaccard coefficient for a value set and a drug class, $J(V, C)$, computes similarity as the ratio between the number of ingredients common to the value set and the drug class, $|V \cap C|$, over the total number of ingredients in the value set and the drug class, $|V \cup C|$. The Jaccard coefficient ranges from 0 (no similarity) to 1 (exact match). The (single or combination) drug class with the highest Jaccard score is considered to best reflect the intension of the value set.

$$JC(V, C) = \frac{|V \cap C|}{|V \cup C|}$$

Figure 2 presents an example, in which the value set *Beta Blocker Therapy* is evaluated against the MeSH class *Adrenergic beta-Antagonists* intersected with the NDF-RT disease class *Hypertension*. The red dotted line indicates that there are 14 ingredients in the combination class. The white circle indicates that 12 ingredients are in both the value set and combination class. This results in a Jaccard score of 0.86 (12/14). The intension for *Beta Blocker Therapy* value set could then be interpreted as *Adrenergic beta-Antagonists* used for *Hypertension*.

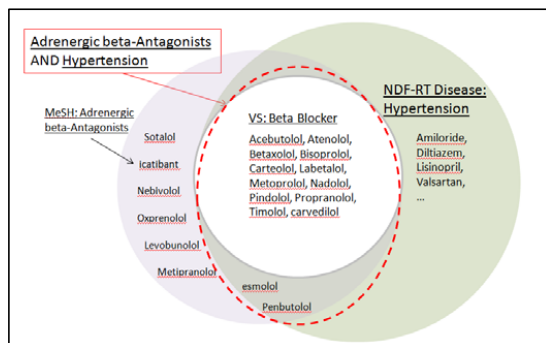


Figure 2 - The Jaccard metric is used to assess the equivalence between value sets and drug classes. The Beta Blocker Therapy value set has an initial match with Adrenergic beta-Antagonists from MeSH. This match is further refined by intersecting the MeSH class with the NDF-RT disease class Hypertension.

Evaluation

We performed a quantitative and qualitative evaluation to assess the fit and validity of elicited intensions.

Quantitative. The quantitative evaluation assesses the overall similarity between value sets and (single or combination) drug classes. More specifically, we simply compute the average of the best Jaccard score for each value set-drug class pair. To assess the contribution of the combination classes (intersection and union classes defined earlier), we compared the averages obtained under the following strategies:

1. When using only single classes
2. When using single classes and intersection classes
3. When using single classes, intersection classes and union classes

We conducted a one-way ANOVA (with repeated measures) to compare the effect of these different strategies on eliciting the intension based on the Jaccard score. A Tukey *post hoc* analysis was performed to identify which strategies were significantly different (pairwise). The statistical analysis was completed using STATA 13 (StataCorp. 2013. College Station, TX).

Qualitative. The qualitative evaluation assesses the extent to which the (single or combination) drug class matches the intension reflected in the value set name. An expert pharmacist (SDN), who had not been involved with the development of the methods, analyzed the drugs listed in the class and in the value set for the class identified as the best match for each value set. More specifically, the pharmacist was asked to answer two main questions for each value set-drug class pair:

1. Do the drugs listed in the value set correspond to the intension reflected in the value set name?
2. Do the drugs listed in the (single or combination) class correspond to the intension reflected in the value set name?

Additionally, the pharmacist was asked whether there were missing or extraneous drugs in the value set, in the best-matching class, or in both.

Results

Establishing sets of ingredients from drug value sets

As shown in Figure 1, there were 183 extensional medication value sets. Ninety-seven were excluded because they contained only one ingredient and could be trivially mapped to the Chemical structure drug class restricted to this ingredient. Five were excluded because their extensions were composed entirely of multi-ingredient drugs. One contained one ingredient that could not be mapped to a drug class. The remaining 80 value sets were analyzed and contained 468 distinct ingredients after mapping to RxNorm.

Establishing sets of ingredients from drug classes

Table 1 shows the number of single and combination classes for each source. The NDF-RT disease classes (reflecting drug indications) were combined with other drug classes resulting in approximately 4 million intersections. We iteratively generated the union of classes that had some equivalence to value sets based on the Jaccard score. This resulted in approximately 100 union candidates. The 6519 single classes contained 2957 distinct ingredients after mapping to RxNorm.

Table 1 - Number of drug classes in each source.

Type	Source	# classes
Single	ATC	882
	MeSH Pharmacological Action	350
	DailyMed Established Pharmacological Class	431
	DailyMed Mechanism of Action	348
	DailyMed Physiological Effect	239
	NDF-RT Disease	1434
	NDF-RT Chemical Structure	2835
Combination	NDF-RT Disease intersected with all single drug classes	Approximately 4 million
	Union of multiple drug classes	Approximately 100

Comparing ingredients between value sets and drug classes

We compared each of the 80 value sets to all single drug classes, resulting in 521,520 comparisons, from which we selected the best value set-drug class pair match. We found that there was no single source that best described all value sets. However, many of the top matches appear to come from ATC and DailyMed EPC, followed by NDF-RT Physiologic effect and NDF-RT Mechanism of action.

Similarly, we identified the best match for each value set and all intersection drug classes (over 320 million comparisons). Here again, intersections between Disease classes and classes from ATC and DailyMed EPC provided most of the best matches. The intersection of Disease classes and classes from NDF-RT Chemical structure also contributed many of the best matches. Finally, we determined the best union classes for each value set.

Evaluation

Quantitative. We examined the distribution of the Jaccard scores for best matches between the 80 value sets and drug

classes, using three mapping strategies: single classes only; single classes or intersection classes; single classes or intersection classes or union classes. A larger proportion of the value sets obtain better Jaccard scores under the last two strategies compared to single classes.

More specifically, the average Jaccard score of single classes was 0.65, with 23% (18/80) of the value sets having high similarity (0.9-1) with a drug class. Adding disease intersections increased the average Jaccard score to 0.79 with 34% (27/80) of the value sets having high similarity with a single drug class or a drug class intersected with a disease class. Adding union classes provided a very small performance gain, only increasing the average Jaccard score to 0.80. The union mostly affected value sets that did not have a good match to a single drug class or single drug class intersected with a disease class.

The one-way ANOVA test showed that the different strategies resulted in statistically significant differences in mean Jaccard score, $F(2, 158) = 30.02$, $p < 0.005$. The *post hoc* Tukey test indicated that both types of combination classes performed significantly better than single classes only ($p = 0.024$ and $p = 0.009$). However, there was no significant difference between the two types of combination classes ($p = 0.934$).

Table 2 - Examples of elicited intensions (original value set, single or combination drug class, Jaccard score, and pharmacist comments).

Value Set → Drug class [Pharmacist comments]	JC
Leukotriene modifiers → Leukotriene Antagonists [No comments, exact match.]	1.00
IV or IM Beta-lactams → Cefotaxime [Value set missing a lot of drugs.]	0.29
IV PO Quinolone Antipneumococcal → Pneumonia, Bacterial, AND Quinolones [Would have been more valid to filter on route.]	1.00
Beta Blocker Therapy for LVSD → Tachycardia, Supraventricular AND Adrenergic beta-1 Receptor Antagonists [Value set only contains oral forms of the drugs and there are differences in the salt forms.]	0.40
Beta-Lactams for Allergy Determination → beta-Lactams OR other beta-lactam antibacterials OR Beta-lactamase sensitive penicillins [Value set had a lot of missing SCDs. Drugs that are not antibiotics, but inhibit beta-lactams not in ingredients.]	0.80
BH Mood Stabilizer Medication RxNorm → Phenothiazines with piperazine structure OR Dibenazepines OR Fructose OR Other antipsychotics [Value set is poorly defined and has a lot of missing drugs.]	0.27

Qualitative. Table 2 shows examples of value sets with elicited intensions.

Value set extension vs. value set name. Eighty percent (64/80) of the value sets had ingredients that were consistent with the intension reflected in the value set name. Sixteen percent (13/80) of the value sets had extraneous ingredients and 46% (37/80) had missing ingredients.

Drug class extension vs. value set name. Fifty-one percent (41/80) of the (single or combination) drug classes selected as the best mapping for a value set matched the intension reflected in that value set name. Twenty-eight percent (22/80) of the best-matching drug classes had extraneous ingredients and 43% (35/80) had missing ingredients. In thirty-one percent (25/80) of the cases, there were missing ingredients in both the drug class and value sets.

Other issues. We found several duplicate value sets, such as *BH Antidepressant Medication - SSRI Antidepressants RxNorm* and *SSRI Antidepressants*. The issue of duplicate

value sets had already been reported by Winnenburger et al [7] for disease and procedure value sets, so it was not surprising to find duplicates in medication value sets. We also found some naming inconsistencies, such as with the qualifiers used to refer to injectable drugs in value set names, which included *IV*, *parenteral*, *Injectable*, and *for IV Administration*.

Discussion

Applications

This work presents a framework for eliciting an explicit form of the intension in reference to a normative drug source, such as the ATC, MeSH, NDF-RT and DailyMed drug classification systems. The drug classes from these sources allow us to compare logical groupings of drugs and derive a reference extension that can be used to validate the ingredients in the value set.

There are two main applications for this framework.

The primary application is to derive operational intensional definitions for medication value sets in order to facilitate their maintenance. Once a value set has been associated with a (single or combination) class, value set developers can rely on the corresponding drug classification systems for the maintenance of the value set. In practice, instead of relying on experts for checking whether drugs should be added to or removed from a given value set, value set developers would only need to check the members of the drug class the value set has been mapped to when new versions of the drug classification system become available. Operational intensional definitions greatly facilitate the maintenance and reliability of drug value sets. We discussed our findings with representatives of the NLM Value Set Authority Center (VSAC), where these techniques could be implemented to facilitate the development and ongoing maintenance of the value sets.

A secondary benefit of comparing value sets to drug classes is that it provides an opportunity for quality assurance. Our expert pharmacist identified potential errors, specifically missing drugs and extraneous drugs, in almost two thirds of the value sets. Here again, the availability of such a framework will likely benefit the quality assurance of medication value sets in VSAC. Of note, similar errors were identified in drug classes and could be reported to the developers of the corresponding drug classification systems.

Finally, this work informed the visualization of drug classes in RxClass (<https://mor.nlm.nih.gov/RxClass/>), the tool we developed for browsing and comparing drug classes associated with RxNorm.

Failure analysis

We were able to elicit an appropriate intension for about half of the value sets, while the other half corresponded to value sets with a mixture of drugs belonging to many different, but similar classes. For example, the value set *IV Quinolones Used For Prophylaxis for Hysterectomy and Colon Surgery* best matched the intersection between two drug classes, *Quinolone Antimicrobial* and *Sinusitis*. Antimicrobial drugs have many therapeutic uses, so in this case, we elicited the wrong intension that provided a good match (treatment of sinusitis vs. prophylaxis for hysterectomy and colon surgery).

In other cases, the value sets contained a complicated mixture of drugs. For example, the *Anti-Hypertensive Pharmacologic Therapy* value set contained 67 distinct ingredients. The best match consisted of a combination class resulting from the intersection of *Hypertension* with the top-level class *Established Pharmacologic Class (EPC)*, i.e., all drugs with any EPC class. Interestingly here, the expected best match was to the single

class *Hypertension* itself, not an intersection class. However, this occurred because several drugs used to treat hypertension were missing from the value set, such as *Methyldopa*, *Guanethidine*, and *Bumetanide*. This case illustrates not so much a failure of our approach to associating value sets with drug classes than a quality issue with the value set.

We observed that some of these value sets were restricted by dose form (e.g., *IV Quinolones*, which restricts antibiotics from the quinolone class to their intra-venous forms). Because drug classes list ingredients as their members, a drug class, such as *QUINOLONE ANTIBACTERIALS* in ATC would retrieve the appropriate ingredients, but would not distinguish between intra-venous and other forms of these ingredients. For these classes, RxNorm could be used to filter ingredients for which there exist clinical drugs for specific dose forms. Failure to implement this feature in our framework resulted in extraneous ingredients in some value sets. For example, *QUINOLONE ANTIBACTERIALS* in ATC lists the ingredient *Ofloxacin* as a member, for which there exist no injectable forms. Similarly, the dose of the drug often had contextual implications. For example, some value sets only listed heparin flushes, rather than the therapeutic dose. This suggests that value set developers created some value sets with specific dose form groups in mind.

Finally, there were a few cases, where we were unable to achieve a good match because the requisite information was often outside the scope of the terminology. For example, certain drugs or doses listed in the terminology are not legally prescribed in the U.S., such as *Phenprocoumon* and *Dicumarol*.

Limitations and future work

There are a few limitations regarding the materials used in this study. First, multi-ingredient drugs were not included in the analysis, because the corresponding single-ingredient drugs tend to be listed in the value sets and also because multi-ingredient drugs are not represented consistently across drug classification systems. However, this restriction resulted in eliminating five value sets composed entirely of multi-ingredient drugs. Second, the therapeutic intent of medications is often dependent on medication route and dose. As discussed in the failure analysis, we could easily restrict ingredients based on intended route, when a route restriction was expressed in the value set name. However, we would probably not be able to account for specific doses (such as heparin flushes), because this information was not made explicit in value set names. Third, this study is based on older versions of the value sets, RxNorm and drug classification systems, because the qualitative analysis was performed against these specific versions. Nevertheless, the methods are generalizable to newer versions.

Conclusion

We proposed an approach for eliciting the intension of medication value sets by comparing the list of ingredients in these value sets to single or combination drug classes derived from drug classification systems, such as ATC, MeSH, NDF-RT and DailyMed. With this approach, we were able to find drug classes that match the value sets with 0.79 Jaccard similarity on average. This approach has the potential for facilitating the development and maintenance of medication value sets. We also discussed its benefits in terms of quality assurance.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine (NLM). During this project, Dr. Nelson was supported by the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, Department of Veterans Affairs. The views, findings, and conclusions expressed in this report are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, the Food and Drug Administration, or the National Library of Medicine.

References

- [1] Meaningful Use Definition and Meaningful Use Objectives of EHRs | Providers & Professionals | HealthIT.gov., Available: <https://www.healthit.gov/providers-professionals/meaningful-use-definition-objectives>, Accessed: December 5 2016
- [2] CMS.gov, Available: <https://www.cms.gov>, Accessed: December 5 2016
- [3] P.C. Tang, M. Ralston, M.F. Arrigotti, L. Qureshi, and J. Graham, Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures, *J Am Med Inform Assoc* **14** (2007), 10-15.
- [4] F. Eisenberg, C. Lasome, A. Advani, R. Martins, P. Craig, and S. Sprenger, A study of the impact of meaningful use clinical quality measures, *American Hospital Association* (2013).
- [5] K.S. Chan, J.B. Fowles, and J.P. Weiner, Review: electronic health records and the reliability and validity of quality measures: a review of the literature, *Med Care Res Rev* **67** (2010), 503-527.
- [6] L.M. Kern, S. Malhotra, Y. Barron, J. Quaresimo, R. Dhopeswarkar, M. Pichardo, A.M. Edwards, and R. Kaushal, Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study, *Ann Intern Med* **158** (2013), 77-83.
- [7] R. Winnenburg and O. Bodenreider, Metrics for assessing the quality of value sets in clinical quality measures, *AMIA Annu Symp Proc* **2013** (2013), 1497-1505.
- [8] RxNorm, Available: <https://www.nlm.nih.gov/research/umls/rxnorm/>, Accessed: December 5 2016
- [9] WHOCC - ATC/DDD Index, Available: https://www.whooc.no/atc_ddd_index/, Accessed: December 5 2016
- [10] Medical Subject Headings, Available: <https://www.nlm.nih.gov/mesh>, Accessed: December 5 2016
- [11] S.H. Brown, P.L. Elkin, S.T. Rosenbloom, C. Husser, B.A. Bauer, M.J. Lincoln, J. Carter, M. Erlbaum, and M.S. Tuttle, VA National Drug File Reference Terminology: a cross-institutional content coverage study, *Stud Health Technol Inform* **107** (2004), 477-481.
- [12] DailyMed, Available: <https://dailymed.nlm.nih.gov/dailymed>, Accessed: December 5 2016

Address for correspondence

Olivier Bodenreider, MD, PhD
8600 Rockville Pike, 38A/09S904, Bethesda, MD 20894
Phone Number: (301) 827-4982
E-mail: olivier.bodenreider@nih.gov

Comparison of Three English-to-Dutch Machine Translations of SNOMED CT Procedures

Ronald Cornet^{a,b}, Carly Hill^a, Nicolette de Keizer^a

^a Department of Medical Informatics, Academic Medical Center – University of Amsterdam, Amsterdam Public Health Research Institute, The Netherlands

^b Department of Biomedical Engineering, Linköping University, Linköping, Sweden

Abstract

Dutch interface terminologies are needed to use SNOMED CT in the Netherlands. Machine translation may support in their creation. The aim of our study is to compare different machine translations of procedures in SNOMED CT. Procedures were translated using Google Translate, Matecat, and Thot. Google Translate and Matecat are tools with large but general translation memories. The translation memory of Thot was trained and tuned with various configurations of a Dutch translation of parts of SNOMED CT, a medical dictionary and parts of the UMLS Metathesaurus. The configuration with the highest BLEU score, representing closeness to human translation, was selected. Similarity was determined between Thot translations and those by Google and Matecat. The validity of translations was assessed through random samples. Google and Matecat translated similarly in 85.4% of the cases and generally better than Thot. Whereas the quality of translations was considered acceptable, machine translations alone are yet insufficient.

Keywords:

SNOMED CT; Natural Language Processing

Introduction

The use of SNOMED CT in electronic health systems is growing [1]. Recording clinical data using SNOMED CT helps to uniformly describe medical data, which enables data reuse such as data analysis, auditing quality of care and decision support.

SNOMED CT is officially released in English and Spanish. Before SNOMED CT can be used in clinical practice in other languages, a translation needs to be made, or interface terminologies need to be created. SNOMED CT has been fully translated in Denmark and in Sweden. A Canadian French translation is ongoing. Similar to other countries [2], in the Netherlands a partial translation of SNOMED CT is undertaken. Such translations generally provide one term for each concept, adhering to the strict translation guidelines of SNOMED International¹. However, they do not necessarily contain the synonyms used in clinical practice, and these synonyms may not adhere to the translation guidelines. Interface terminologies provide a close-to-user description for concepts, generally covering a part of SNOMED CT, e.g., diagnoses or procedures. Also for Dutch healthcare, one or more Dutch interface terminologies for SNOMED CT need to be made available. Descriptions for a part of the diagnoses form the Dutch interface terminology "Diagnoses thesaurus",

which is maintained by Dutch Hospital Data (DHD) ². The next step is to start creation of an interface terminology for procedures. Recording of procedures is an essential part of clinical documentation and serves, once standardized, many data reuse purposes. These include calculation of quality indicators and reimbursement.

SNOMED CT contains more than 55,000 procedures, each described by one or more English descriptions. Manual translation of these descriptions requires a lot of time and resources. If computers are used to make initial translations, terms only have to be validated, which may save a lot of time. Machine translation has already been used for translating SNOMED CT in Spanish, Swedish and French. To make the Spanish version, prefixes, suffixes and roots of terms were used to make an automated proposal [3]. In Sweden, mappings to other already translated terminologies were used [4]. In France, lexical methods and mapping to the Unified Medical Language System Metathesaurus (UMLS Metathesaurus) were used [5].

Above methods were well evaluated, but not much research has been performed on already available translation tools. Hence, in this study, we assess the quality of the translation of descriptions of concepts in the procedures hierarchy of SNOMED CT from English to Dutch. We compare generic translation tools and a tool with a translation memory that was specifically trained and tuned for this purpose.

The first generic tool is Google Translate³, the second generic tool is Matecat⁴ [6], and the third tool is Thot⁵ [7], a toolkit for statistical machine translation, which requires training and tuning of the translation memory.

The hypothesis is that translations that are the same among the three methods are of better quality, because three different translation systems have found the same result. We furthermore hypothesize to get less but better translations with Thot as this tool is trained with terms from a medical background.

Materials & Methods

Tools

Machine translation was performed using Google Translate, Matecat, and Thot.

Google Translate is a widely used translation tool. The translator is a black box with a large translation memory. It

¹ <http://www.ihtsdo.org/resource/resource/9>

² <https://www.dhd.nl/klanten/producten-diensten/diagnosethesaurus/Paginas/Diagnosethesaurus.aspx>

³ <https://translate.google.com/>

⁴ <https://www.matecat.com/>

⁵ <http://daormar.github.io/thot/>

has previously been used for translation of SNOMED CT, e.g., in the research of Schulz et al. [2].

Matecat [6], is a tool that has its own translation memory. One difference with Google Translate is that one can select the subject of the text to translate. This context might help with choosing the right translation for a term. We selected ‘Medical/Pharmaceutical’ as the subject of the Matecat translation.

Thot is a phrase-based tool [7], i.e., it is usually trained by providing pairs of translations of phrases. In this research, these pairs consisted of Dutch and English phrases, predominantly noun phrases. Apart from providing the basic functionality for preparing, training and translating files, this tool also provides interactive translation possibilities. This means the system provides functionality for the user to have an influence on the final translations. For training, incremental learning is used to result in better language models.

Sources for Training, Tuning and Testing Thot

We used three resources to train, tune and test Thot. First, we used Dutch-English translation pairs from the UMLS Metathesaurus. Second, an existing but partial Dutch translation of SNOMED CT that has been developed in the joint Dutch-Belgian efforts to develop Dutch interface terminologies. Third, a Dutch-English medical dictionary, Springer Groot Medisch Woordenboek [8].

Thot tests were compared by means of the BLEU (bilingual evaluation understudy) score generated by Thot [9]. This score measures the closeness to a human translation in a range from zero to one, with one being a perfect human translation. The configuration that gave the best test result after training and tuning was used for the final translation of the SNOMED CT descriptions. This resulted in the Thot translation, which was compared to the Google Translate and Matecat translations.

Filtering non-translations

The fully specified names of all procedure concepts of the January 2016 release of SNOMED CT were selected. Each tool created a translation for each of the fully specified names, after removal of the semantic tags “(procedure)” and “(regime/therapy)”.

Tools may fail to translate some words, resulting in untranslated English words in a “Dutch” translation. For example, if “Urinary undiversion” is translated as “urine undiversion,” the second word is not a Dutch word. However, if “Open drainage of liver” is translated as “Open drainage van de lever,” this is perfectly correct.

Hence, before comparing the three translations the amount of remaining English words was assessed. This was done by a Java program that checked if the words were in a Dutch list of words. This list consisted of generic Dutch words⁶ merged with the words from the Dutch file we used to train Thot for domain-specific words. Translations containing more non-Dutch words than Dutch words were deleted. The amounts and percentages of included terms were calculated for each of the three tools, and the results were compared with McNemar tests.

Validity of translations

For each SNOMED CT concept we constructed a set of translations, as shown in Table 1. Sets with three equal translations, two equal translations and all different translations were created. We calculated the percentage of sets with at least one exactly similar translation by another method.

From each of the sets, a sample of 100 English terms was selected, with one, two and three different Dutch translations respectively. These samples of the translations were checked on validity: two reviewers (RC & CH) assessed whether the translations were well-formed Dutch noun-phrases reflecting the meaning of the English description. The meaning of the translation could be different, for example due to translation of an English homonym in the wrong context, such as vessels as ships (“schepen”) (see Table 1), or stool as furniture (“kruk”). The reviewers graded using marks from zero to three, with zero and one being not acceptable and two and three being acceptable as a good translation. If an English word is recognizable as a Dutch word, the term will get 2 points, otherwise not more than 1 point. Determiners are not considered; wrong spacing costs a point, as Dutch is a language in which words are combined, e.g., “Mouth reconstruction” is “mondreconstructie”, not “mond reconstructie”.

A translation was considered acceptable when it was recognizable as a translation that covers the meaning of the English term. Average marks and 95% confidence intervals (CI) were calculated for the different samples and reviewers. Also, total weighted averages were calculated for the three tools. The percentages of the translations that were considered acceptable were calculated. For the samples containing two or three different translations, the reviewers assessed which translation they considered the best.

Table 1 – Examples of three sets of translations made by the tools.

	English term	Google	Matecat	Thot
1	Oral sedation	orale sedatie	orale sedatie	orale sedatie
2	CT of pancreas	CT van de alveesklier	CT van de alveesklier	ct pancreas
3	Ultrasound scan of abdominal vessels	Echografie van abdominale schepen	Echografie van de buik vaartuigen	echografie abdominale bloedvaten

Results

Training Thot

Table 2 shows the results of the different Thot translations with their BLEU scores. The configuration with the highest score was used to compare to the general translation tools. This turned out to be a combination of Dutch terms (coming from other thesauri like ICD, ICPC) that UMLS relates to SNOMED CT concepts, terms from a Dutch-English medical dictionary, and terms from a Dutch translation of parts of SNOMED CT.

Filtering non-translations

SNOMED CT contained 54419 procedure concepts.

Sets of translations (i.e., all translations for a concept) were excluded if any of their translation contained more English words than Dutch words. Table 3 shows the number and percentage of included translations for each of the tools used.

⁶ <http://www.opentaal.org/bestanden.html>

Table 2 – BLEU scores generated by Thot for different training and tuning configurations. UMLS CT = SNOMED CT concepts from UMLS; UMLS procedures = SNOMED CT procedures from UMLS; Dict = Medical Dictionary; Trans = partial translation of SNOMED CT in Dutch.

Terms in training and tuning files	BLEU score
UMLS CT + Dict + Trans (tuned with procedures)	0.596
UMLS CT + Dict (tuned with procedures)	0.430
UMLS CT (tuned with procedures)	0.427
UMLS procedures	0.407
UMLS CT	0.357

Table 3 – Translation with amount and percentage of included terms. Total amount of terms to translate was 54419.

Translation	Number and percentage of translations included
Google	52399 (96.3%)
Matecat	52414 (96.3%)
Thot	52685 (96.8%)

Table 4 – Amount of translations after checking and comparing files. Total number of included terms was 50838.

Equal translations	Number (percentage) of terms
All translations equal	1548 (3.0%)
Two different translations	42132 (82.9%)
- Google & Matecat vs. Thot	41865
- Google & Thot vs. Matecat	180
- Matecat & Thot vs. Google	87
All translations different	7158 (14.1%)

There was no practical difference in the number of translated terms, whereas Thot translated ($p < 0.001$) significantly more terms into Dutch than Google and Matecat, which were not significantly different ($p = 0.535$). In all translations, less than 4% of the translations were rejected. Excluding all preferred

terms for which one or more of the tools didn't provide a translation resulted in a set of 50838 terms. This set was used for further analysis.

For these terms we compared whether one or more of the tools provide the same translation. The results of this analysis are shown in Table 4.

Table 4 shows that full agreement between the three tools is much less common than full disagreement, and that Google and Matecat agree most of the time, in 43413 (85.4%) of the cases.

Validity of translations

We selected 100 sets from 1548 with all translations equal, i.e., 100 translated terms; 100 sets from 41865 where Google and Matecat agreed, but Thot did not, hence 200 translated terms, and 100 sets from 7158 that had 3 different translations each, hence 300 translated terms.

Table 5 presents the results of the analysis on the mean acceptability score of each translation, the percentage of acceptable translations, and the agreement between the two raters on this judgement.

The total mean translation score for Google Translate was 2.15. The total mean translation score of Matecat was 2.11 and of Thot 1.91. Taking 0 and 1 (regarded as not acceptable) and 2 and 3 (regarded as acceptable) together resulted in acceptability percentages between 45% and 93%.

Finally, in those cases where there was more than one translation for a term, the reviewers determined which tool they considered to provide the best translation. This is shown in Table 6.

Kappa's for acceptability were 0.623 and 0.577, meaning a fair level of agreement. Both were significant with $p < 0.001$. The weighted average of the percentages from Table 5, based on the number of terms from Table 4, results in an overall percentage of acceptable terms of 61%. The translations by Matecat and Google Translate were considered better than those of Thot.

Table 5 – Mean translation scores per reviewer, percentage of translations considered acceptable by both reviewers, kappa for acceptability and p-value of kappa.

Translation	Mean score (95% CI)	Acceptability Percentage acceptable	Kappa (p-value)
Equal translations		93%	0.712 ($p < 0.001$)
- Reviewer 1	2.6 (2.5 - 2.8)		
- Reviewer 2	2.7 (2.6 - 2.9)		
Two different; translation Google & Matecat		67%	0.400 ($p < 0.001$)
- Reviewer 1	2.1 (1.9 - 2.3)		
- Reviewer 2	2.1 (1.9 - 2.3)		
Two different; translation Thot		53%	0.523 ($p < 0.001$)
- Reviewer 1	2.0 (1.8 - 2.1)		
- Reviewer 2	1.8 (1.7 - 2.0)		
All different; translation Google		81%	0.277 ($p = 0.002$)
- Reviewer 1	2.4 (2.3 - 2.5)		
- Reviewer 2	2.3 (2.1 - 2.4)		
All different; translation Matecat		64%	0.421 ($p < 0.001$)
- Reviewer 1	2.1 (2.0 - 2.3)		
- Reviewer 2	2.0 (1.8 - 2.2)		
All different; translation Thot		45%	0.401 ($p < 0.001$)
- Reviewer 1	1.9 (1.7 - 2.1)		
- Reviewer 2	1.7 (1.6 - 1.9)		

Table 6 – Percentage of translation scored as best by both reviewers.

Translation	Percentage regarded best
Two different: Google & Matecat	42%
Two different: Thot	38%
Two different: no agreement	20%
All different: Google	39%
All different: Matecat	17%
All different: Thot	17%
All different: no agreement	27%

Discussion

In this research the quality of translations by three machine translation engines was tested. To our knowledge, this is among the first studies in which different automated translations were compared.

The Thot translation improved when it was tuned with procedures. The big step to making a translation that resembled human translation was training with the available parts of Dutch SNOMED CT. This training set already contained some translations for terms that had to be translated by Thot.

Even though Thot used existing translations for SNOMED CT descriptions, it could not outperform the tools with a generic translation memory.

Translations that are the same among the three methods are generally of better quality. Ninety-three percent of those translations are acceptable, a number which is only approached by the Google translations in the case that all tools provide different translations. We expected to get less but better translations using Thot, as it was trained for this purpose. However, it has a much smaller translation memory than Google and Matecat, and the contrary has been the case. The number of translations, after filtering non-translations, was higher for Thot than for Google and Matecat, but the validity of the translations was lower.

A Java program checked the translations for non-Dutch words, and we were surprised by the large amount of terms that could be translated. This does not directly mean the translations are of good quality, but it does mean that most terms have been translated to Dutch terms.

In the comparison Matecat and Google translated many terms the same, and Thot translated very differently. This emphasizes the difference between generic and specific tools. The real quality was measured by manually assessing the validity of three samples of 100 terms. Average scores show the terms are on average considered acceptable. However, we did see that the Thot translations were the only ones to score under 2.00 on average, and under 60% in acceptable translations. This means the Thot translations were considered inferior to the other translations. There is also a difference in the total average score. The terms that were all translated the same were considered better than the other terms. The reviewers could see that these terms were shorter than other terms. This might mean there was less possibility for making different or wrong translations. When the two different translations were compared, Matecat and Google scored a bit better than Thot. When comparing the three different translations, the Google translation was considered the best most of the time. Most of the translations were considered acceptable, but not perfect. The translations made by Google, Matecat and Thot can give an idea of the translation of a term, but most of the time they will not give a translation that is

ready to use. In this research, Google Translate scored the best in comparison with Matecat and Thot. Thot scored worst.

Limitations of this study were the fact that only two reviewers rated the terms, and the small size of the samples that were judged for validation of the translations. Furthermore, the analysis of acceptability of translations could be further systematized. Additionally, the types of errors in translations could be further specified. For example, the generic tools occasionally used synonyms from a non-medical domain, leading for example to translations of “blood vessels” as “blood ships.” Regarding Thot, the high inclusion rate (i.e., relatively few English terms in a translation) seems to be explained by the fact that Thot simply leaves out fragments it cannot adequately translate. This may lead to close-to-human translation, but lacks the full semantics of the English term that has to be expressed. Further analysis of such mechanisms and types of error is needed.

Creating translations with machine translation tools has the potential to help make good translations, but the translations are not made according to the translation rules for SNOMED CT from SNOMED International. This means making a translation with machine translation tools will only provide reference terms, and not official translations. This makes machine translation not suitable for making an official translation of SNOMED CT, as stated in the research of Schulz [2].

Our research contributes to the field of machine translation, in which efforts are undertaken for various languages. The distinguishing features are the application to SNOMED CT, and the translation to Dutch. In earlier research [10], we showed that Dutch is among the languages with a scarcity of resources for language processing in medicine. It similarly lacks corpora, especially bilingual medical corpora, that could contribute to improving the quality of machine translation. This poses challenges and gives need to the use of alternative approaches, such as lexical and morphosemantic approaches [11].

Our study is similar to the one described in [12], which compared three approaches for translating the Gene Ontology from English to German. They used Wikipedia, Google API with context and Google API without context. The average scores for adequacy (the extent to which a translation represents the meaning) and for fluency (the extent to which a translation is proper German) were over 4.0 on a 5-point Likert scale. This seems higher than the scores we found in our study, which may be explained by the fact that we combined adequacy and fluency in our scores, and the existence of a larger corpus of German terms, as German is the second-most represented language in PubMed/MEDLINE, after English [12].

Further research should be performed on using machine translation tools in a full translation process. This could determine whether using machine translation tools will be beneficial for translation. Training Thot using only validated translations that conform to SNOMED International’s translation guidelines may eventually lead to higher translation quality. Research could be done on building a good translation memory to train a tool such as Thot. This might result in translations of better quality. Furthermore, this may prove useful for maintenance, i.e., for providing translations for concepts that are added to new releases of SNOMED CT.

Further research is also needed to point out the impact that translations, translation quality, and adequacy of synonymy have when actually using a translation of SNOMED CT in clinical practice, for example, on impacting the inter-coder agreement, which is considered to be low when using SNOMED CT in English [13].

Conclusion

Overall quality of the three different tools was considered acceptable, but not good enough for use in clinical practice. The Thot translations were considered worse than the Google and Matecat translations. Shorter terms were more often translated the same by the three tools, and these translations were considered better. The translations made by the tools could be used in a translation process, but cannot be used directly. The translation tools cannot translate the terms according to the translation rules for SNOMED CT. This means the tools are of limited help for making an official translation of SNOMED CT.

References

- [1] D. Lee, N. de Keizer, F. Lau, and R. Cornet, Literature review of SNOMED CT use, *J Am Med Inform Assoc* **21** (2014), e11-19.
- [2] S. Schulz, J. Bernhardt-Melischnig, M. Kreuzthaler, P. Daumke, and M. Boeker, Machine vs. human translation of SNOMED CT terms, *Stud Health Technol Inform* **192** (2013), 581-584.
- [3] G.A. Reynoso, A.D. March, C.M. Berra, R.P. Strobietto, M. Barani, M. Iubatti, M.P. Chiaradio, D. Serebrisky, A. Kahn, O.A. Vaccarezza, J.L. Leguiza, M. Ceitlin, D.A. Luna, F.G. Bernaldo de Quiros, M.I. Otegui, M.C. Puga, and M. Vallejos, Development of the Spanish version of the Systematized Nomenclature of Medicine: methodology and main issues, *Proc AMIA Symp* (2000), 694-698.
- [4] G.O. Klein and R. Chen, Translation of SNOMED CT - strategies and description of a pilot project, *Stud Health Technol Inform* **146** (2009), 673-677.
- [5] L. Deleger, T. Merabti, T. Lecrocq, M. Joubert, P. Zweigenbaum, and S. Darmoni, A twofold strategy for translating a medical terminology into French, *AMIA Annu Symp Proc* **2010** (2010), 152-156.
- [6] M. Federico, N. Bertoldi, M. Cettolo, M. Negri, M. Turchi, M. Trombetti, A. Cattelani, A. Farina, D. Lupinetti, A. Martinez, A. Massidda, H. Schwenk, L. Barrault, F. Blain, P. Koehn, C. Buck, and U. Germann, The Matecat Tool, *Proceedings of the 25th International Conference on Computational Linguistics (COLING)* (2014), 129-132.
- [7] D. Ortiz-Martinez, and F. Casacuberta, The New THOT Toolkit for Fully-Automatic and Interactive Statistical Machine Translation, *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics* (2014), 45-48.
- [8] P. Reuter, Springer Groot medisch woordenboek. Medical Dictionary. Engels-Nederlands English-Dutch. Springer Uitgeverij Houten (2008) 738 pp. (+ CD rom)
- [9] K. Papineni, S. Roukos, T. Ward, and W-J. Zhu, BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th annual meeting on association for computational linguistics - ACL '02* (2002), 311-318.
- [10] R. Cornet, A. Van Eldik, and N. De Keizer, Inventory of tools for Dutch clinical language processing, *Stud Health Technol Inform* **180** (2012), 245-249.
- [11] O. Perez-de-Vinaspre, and M. Oronoz, Translating SNOMED CT Terminology into a Minor Language, *Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis, Louhi* (2014): 38-45.
- [12] N.D. Hailu, K.B. Cohen, and L.E. Hunter, Ontology translation: a case study on translating the gene ontology from English to German, *International Conference on Applications of Natural Language to Data Bases/Information Systems* (2014), 33-38.
- [13] J.E. Andrews, R.L. Richesson, and J. Krischer, Variation of SNOMED CT coding of clinical research concepts among coding experts, *J Am Med Inform Assoc* **14** (2007), 497-506.

Address for correspondence

Ronald Cornet
 Academic Medical Center – University of Amsterdam
 Department of Medical Informatics
 Amsterdam Public Health research institute
 P.O. Box 22700
 1100 DE Amsterdam
 The Netherlands
r.cornet@amc.uva.nl

Trends in Fetal Medicine: A 10-Year Bibliometric Analysis of Prenatal Diagnosis

Ferdinand Dhombres, Olivier Bodenreider

National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Abstract

The objective is to automatically identify trends in Fetal Medicine over the past 10 years through a bibliometric analysis of articles published in Prenatal Diagnosis, using text mining techniques. We processed 2,423 full-text articles published in Prenatal Diagnosis between 2006 and 2015. We extracted salient terms, calculated their frequencies over time, and established evolution profiles for terms, from which we derived falling, stable, and rising trends. We identified 618 terms with a falling trend, 2,142 stable terms, and 839 terms with a rising trend. Terms with increasing frequencies include those related to statistics and medical study design. The most recent of these terms reflect the new opportunities of next-generation sequencing. Many terms related to cytogenetics exhibit a falling trend. A bibliometric analysis based on text mining effectively supports identification of trends over time. This scalable approach is complementary to analyses based on metadata or expert opinion.

Keywords:

Bibliometrics, Prenatal Diagnosis

Introduction

The availability of new genomic analysis techniques is transforming research and practice in Medicine. This is especially true of Fetal Medicine with the emergence of non-invasive prenatal testing (NIPT) procedures enabled by sequencing of circulating cell-free DNA (cfDNA) from a simple maternal blood sample [14]. This evolution is expected to be reflected through manuscripts published in Prenatal Diagnosis, the official journal of International Society for Prenatal Diagnosis. In fact, such advances in Fetal Medicine are regularly screened by members of the editorial board and summarized in a yearly editorial “In case you missed it” [2,6,7]. For example, cfDNA was discussed in the editorial presenting trends of the year 2015 [6].

Trend analysis often relies on manual review and expert opinion. For example, significant trends have been identified in medical literature, including increase in frequency and complexity of statistical reporting [1] and increase in computerized tomography and magnetic resonance imaging in Radiology research [12]. Bibliometric techniques have also proved useful for identifying trends in scientific disciplines [15,21], and could be used for capturing an unbiased evolution of major themes in Fetal Medicine over a longer period of time. In context of trend analysis, bibliometric techniques of choice are not citation metrics [4,20,26] (e.g., impact factor and h-index), rather those techniques used for analyzing metadata associated with scientific articles [9,11,13,16] (e.g., indexing terms) and the text of these articles [10]. Surprisingly, use of text mining techniques on full-text articles has not been reported for trend analysis purposes.

The objective of this investigation is to automatically identify trends in Fetal Medicine over the past 10 years through a bibliometric analysis of articles published in Prenatal Diagnosis, using text-mining techniques.

Methods

We conducted a bibliometric analysis of 2,423 full-text articles published in Prenatal Diagnosis over a 10-year period, from January 1, 2006 to December 31, 2015. Our approach can be summarized as follows. We extracted salient terms from the articles; calculated their frequencies over time; and established evolution profiles for most frequent terms, from which we derived falling, stable, and rising trends.

Extracting salient Fetal Medicine terms

We processed the full-text articles to extract all sequences of consecutive words (“N-grams”) of 5 words or less, most likely corresponding to medical terms. Let us consider the sentence “Currently, commercial applications of cell-free fetal DNA testing include RhD blood group typing” [19]. Examples of N-grams extracted from this sentence include “fetal” and “DNA” (N=1); “fetal DNA” and “testing include” (N=2); “cell-free fetal DNA” (N=3); “RhD blood group typing” (N=4); and “testing include RhD blood group” (N=5). Not all N-grams are expected to correspond to medical terms, let alone to salient Fetal Medicine terms. We used Apache Solr (<http://lucene.apache.org/solr/>) to extract N-grams.

Intuitively, common English words (i.e., non-medical words) or expressions and general medical terms are unlikely to be terms of interest. In contrast, terms frequently occurring in Prenatal Diagnosis are more likely to be salient terms. Therefore, as shown in Figure 1 we filtered out all N-grams entirely composed of common English words, e.g., “commercial applications” (filter #1); selected N-grams present in more than 10 articles in at least one year (filter #2); selected N-grams present in UMLS Metathesaurus [3], a large medical dictionary (filter #3); but excluded N-grams corresponding to general medical terms (isolated adjectives and terms categorized as “Concepts & Ideas” in UMLS Semantic Network), e.g. “mmol” and “arterial” (filter #4). Finally, one author (FD) manually reviewed terms excluded by these filters and rescued salient Fetal Medicine terms that were not covered by the medical dictionary (e.g., “cell-free fetal DNA”).

Calculating term frequencies

For each medical term, we recorded the number of articles in which it appears, for each year of the decade 2006–2015, and for the whole decade. Additionally, we determined the cumulative proportion of occurrences for each term in each year.

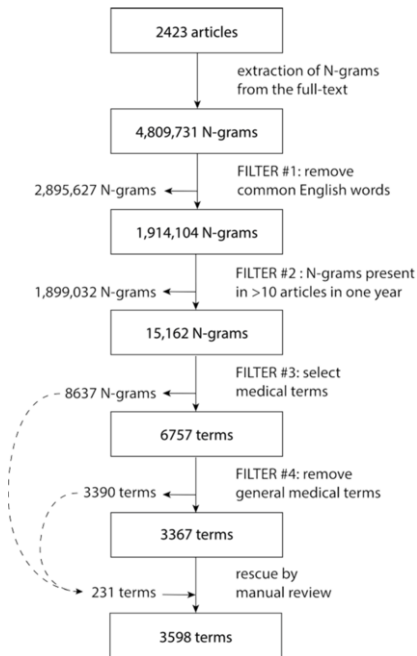


Figure 1 - Term extraction strategy applied to the 2,423 articles from Prenatal Diagnosis (2006-2015). Counts represent numbers of distinct terms (N-grams).

Establishing evolution profiles

Intuition here is that terms used mostly at the beginning of the 10-year period under investigation are becoming less popular (denoting a falling trend). In contrast, terms used mostly at the end of the decade have become more popular recently (denoting a rising trend). In practice, we used cumulative frequency over time to determine when terms were used most. We divided the decade into 3 periods, namely 2006- 2009, 2010-2011, and 2012-2015. For a given term, if 50% or more of all occurrences were observed during 2006-2009, its overall frequency is decreasing (falling trend). In contrast, if 50% or more of all occurrences were observed during 2012- 2015, its overall frequency is increasing (rising trend). Otherwise, the term was deemed “stable”. We extracted 30 most frequent terms in each trend group for visualization and further analysis. Additionally, we extracted 200 most recent terms among those exhibiting a rising trend, as they are likely to denote “hot terms”. Finally, we surveyed frequency evolution for a selection of terms, including those identified by editors of *Prenatal Diagnosis* as reflecting advances in Fetal Medicine for 2015[6].

To produce the evolution profiles, we used the R Foundation Computing environment [17] along with packages for text and data management [24,25] and for data visualization [22,23]. Excluding manual review of terms, it took about four hours to process the documents and compute evolution profiles.

Results

Extracting salient Fetal Medicine terms

From the 2,423 articles, we identified 3,598 salient medical terms. On average, the terms occurred in 101.9 articles over the decade. Our manual review rescued 231 (2.7%) of the 8,637 terms that had been inappropriately filtered out, including

“prenatal ultrasound”, “maternal plasma”, “fetal nuchal translucency” and “cell-free DNA”. These terms were present in 178.5 articles on average, ranging from 33 (for “fetoscopic laser photocoagulation”) to 883 articles (for “fetal medicine”).

Establishing evolution profiles

Distribution of terms according to year in which their cumulative frequency reaches 50% of their total document frequency is presented in Figure 2. We identified 618 terms with decreasing frequencies over time (falling trend), 2,142 stable terms, and 839 terms with increasing frequencies (rising trend). Not surprisingly, while stable terms occur in a large number of articles, terms with decreasing or increasing frequencies occur in fewer articles.

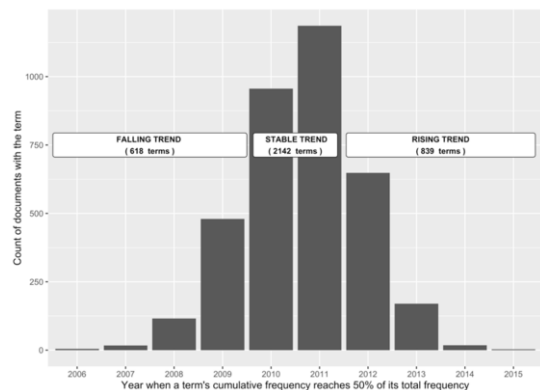


Figure 2 - Distribution of terms according to the year in which their cumulative frequency reaches 50% of their total document frequency.

Falling trend. Among the most frequent terms with decreasing frequencies, we found many terms related to Cytogenetics (e.g., “FISH”, “cytogenetic analysis”, “molecular cytogenetic”, “cytogenetic studies”). Of note, the term “case report” is the term whose frequency decreased most dramatically, dropping from 121 articles in 2006 to 58 articles in 2015. The top 30 terms exhibiting a falling trend are shown in Figure 3a. These terms reached 50% of their total document frequency before 2010.

Stable trend. Not surprisingly, many common terms in Fetal Medicine have relatively stable frequencies (Figure 3b). For example, the terms “pregnancy”, “fetus”, “ultrasound” were present in over 2,000 articles, and the terms “gestational age”, “karyotype”, “maternal age” and “amniocentesis” in over 1,000 articles. The terms “chorionic villus sampling” and “placenta”, present in over 500 articles are also stable over the decade. These terms reached 50% of their total document frequency in 2010 or 2011.

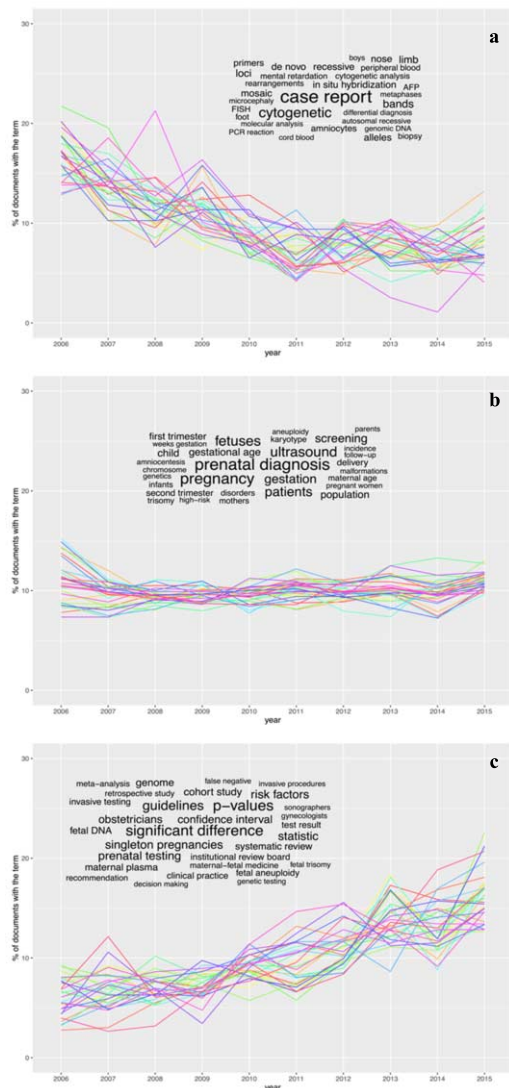


Figure 3 - Evolution of term frequency (coloured lines) over time for the top 30 terms exhibiting a falling trend (a), a stable trend (b) and a rising trend (c). (The font size in the term cloud is proportional to term frequency.)

Table 1 - List of the 20 most frequent of most recent terms exhibiting a rising trend (“hot terms”).

Rank	“hot topic”	Rank	“hot topic”
1	systematic review	11	positive predictive value
2	DNA sequences	12	maternal plasma DNA
3	fetal cell-free DNA	13	non-invasive prenatal diagnosis
4	invasive prenatal testing	14	microarray analysis
5	cell-free DNA	15	clinical setting
6	web	16	non-invasive prenatal testing
7	plasma DNA	17	single nucleotide polymorphism
8	aneuploidy detection	18	prospective cohort study
9	Genomics	19	collaborative study
10	exclusion criteria	20	genetic counselors

Table 2 - Trend for terms in four categories of interest (the most recent terms exhibiting a rising trend are marked ***; the arrows represent rising (↗), stable (→) and falling (↘) trends; df: document frequency).

Category	Term	Df	Trend
invasive diagnostic procedures	amniocentesis	981	→
	chorionic villus sampling	573	→
	fetal blood sampling	128	→
	cordocentesis	116	→
	ultrasound guidance	94	→
next-generation genetics	fetal cell-free DNA ***	244	↗
	fetal cells	229	→
	cell-free DNA ***	210	↗
	microdeletion	177	↗
	comparative genomic hybridization	172	→
	CGH	163	→
	massively parallel sequencing ***	107	↗
	copy number variation	89	↗
	genome sequencing ***	81	↗
	chromosomal microarray ***	79	↗
	CNVs ***	68	↗
	copy number variants ***	65	↗
	whole genome sequencing ***	65	↗
imaging procedures	shotgun sequencing ***	57	↗
	direct sequencing	48	↘
	exome ***	29	↗
	ultrasound	2032	→
	ultrasound examination	683	→
	ultrasound scan	501	→
	ultrasound screening	411	→
	MRI	346	→
	first trimester ultrasound	305	→
	magnetic resonance imaging	299	→
	fetal echocardiography	206	→
	Doppler ultrasound	194	→
	imaging procedures	170	→
fetal therapy procedures	fetal MRI	162	→
	three dimensional ultrasound	160	→
	second trimester ultrasound	102	↘
	X-ray	100	↗
	umbilical artery Doppler	70	→
	transvaginal ultrasound	55	→
	computed tomography	28	→
	heart ultrasound	138	→
	fetal therapy	91	→
	fetal surgery	80	↗
	laser surgery	76	→
	fetal intervention	75	→
	amnioreduction	61	→
	fetoscopy	40	↗
	fetoscopic laser surgery ***	37	↗
	fetoscopic laser coagulation	33	→
	in utero treatment	29	→
	in utero therapy	981	→

Rising trend. Terms with increasing frequencies include those related to statistics (e.g., “statistical analysis”, “p-value”, “significant difference”), medical study design and methods (e.g., “cohort study”, “systematic review”, “meta-analysis”, “ethics committee”, “institutional review board”), and clinical practice documents (e.g., “guidelines”, “recommendations”) (Figure 3c). These terms reached 50% of their total document frequency after 2011. Of particular interest, “hot terms” (i.e., the most recent terms exhibiting a rising trend) generally reflect the new opportunities of next-generation sequencing (“cell-free DNA”, “non-invasive prenatal testing”, “microarray analysis”). The list of the 20 most frequent hot terms is provided in Table 1; these terms exhibit a document frequency ranging from 377 to 123 over the decade.

Trend for specific terms. As expected, many terms identified by editors of Prenatal Diagnosis as reflecting advances in Fetal

Medicine for 2015 [6] were also captured by our approach among the recent terms exhibiting a rising trend (e.g., “fetoscopic laser surgery”, “monochorionic diamniotic twin pregnancies”, “placental insufficiency”, “placental function”). Further analysis of these terms is presented in the discussion section. We also surveyed specific terms for invasive diagnostic procedures, imaging procedures, fetal therapy procedures and next-generation genetics. As shown in Table 2, terms in two of these categories, namely invasive diagnostic procedures and imaging procedures, are generally stable. In contrast fetal therapy procedures and next-generation genetics tend to exhibit a rising trend, some of these terms having appeared very recently (“hot terms”).

Discussion

Trends in Fetal Medicine

Through a bibliometric analysis of articles published in *Prenatal Diagnosis*, using text mining techniques, we were able to identify trends in Fetal Medicine over the past 10 years.

Trends for diagnostic techniques. As expected, terms related to noninvasive prenatal testing exhibit a rising trend. More generally, terms denoting new genetic methods (e.g., “next generation sequencing”, “whole genome sequencing”, “single nucleotide polymorphism” or “microarray analysis”) are on the rise. In contrast, terms related to Cytogenetics (e.g., “molecular cytogenetic” or “FISH”) were highly used at the beginning of the decade, but are now less popular, reflecting a paradigm shift in Fetal Medicine. Interestingly, terms denoting invasive sampling techniques (“amniocentesis”, “choriocentesis”) remain stable in Fetal Medicine discourse, with a high number of occurrences across the decade, possibly because they continue to be mentioned as a reference when discussing newer techniques.

Trends in study design. In addition to trends for diagnostic techniques, our analysis identified trends in study design, namely an evolution toward structured studies reflected by a falling trend for “case report”, as well as a rising trend for “retrospective study” and for “meta-analysis”. The most recent terms (“hot terms”) include “prospective cohort study” and “systematic review”. Case reports are still given consideration for publication as research letters in *Prenatal Diagnosis*. However, a partnership with the journal *Clinical Case Reports* since 2013 may be the reason why fewer case reports end up being published in *Prenatal Diagnosis* nowadays. The rising trend for statistical methods, tests and variables is consistent with the observed evolution of study design towards structured epidemiological and clinical studies reported in the general medical literature [1].

Text mining vs. expert opinion

While our analysis is generally consistent with the trends identified by the editors of *Prenatal Diagnosis* as reflecting advances in Fetal Medicine over the past few years [2,6,7], some terms related to fetal surgery do not appear in our lists of terms exhibiting a rising trend, simply because their frequency is below that of top terms in this group. For example, although it exhibits a rising trend, the term “fetoscopic laser surgery” occurs only in 40 articles during the decade. Similarly, the terms “fetal therapy”, “in utero treatment”, “fetoscopy”, “fetal surgery”, “diaphragmatic hernia”, “spina bifida” or “twin-twin transfusion syndrome” are stable but occur in less than 210 articles. Evaluation of placental function was also deemed as a major advance in 2015 [6], and our analysis also finds a rising trend (but

limited frequencies) for “placental function”, “placental dysfunction” and “placental insufficiency”. Interestingly, although clearly identified in our analysis, trends in study design discussed above were not reported in editorials of the journal (probably because they do not reflect advances in diagnostic techniques per se). Moreover, stable and falling trends are not reported in editorials, but they are identified by our bibliometric analysis.

Text mining vs. metadata analysis

The medical literature referenced in PubMed/MEDLINE is indexed with Medical Subject Heading (MeSH) thesaurus. Therefore, an analysis of the indexing terms (MeSH descriptors) assigned to *Prenatal Diagnosis* articles could also help identify trends in Fetal Medicine. However, MeSH descriptors have limited granularity and there is often a delay between publication and indexing.

Limited granularity. MeSH has a limited number of descriptors for indexing *Prenatal Diagnosis* articles. In addition to the descriptor “*Prenatal Diagnosis*”, there are 7 more specific descriptors, namely “*Amniocentesis*”, “*Chorionic Villi Sampling*”, “*Fetoscopy*”, “*Maternal Serum Screening Tests*”, “*Ultrasonography, Prenatal*”, “*Cervical Length Measurement*”, and “*Nuchal Translucency Measurement*”. Arguably, this granularity is insufficient for specific bibliometric analyses and cannot match granularity resulting from text mining techniques.

Delay between publication and indexing. There is a delay between time of publication and indexing. For example, in May 2016, 53% of articles published by *Prenatal Diagnosis* in 2015 were still awaiting indexing. Moreover, MeSH thesaurus is updated on a yearly basis, with some exceptions for public health emergencies (e.g., the term “*Zika Virus Infection*” was added to MeSH ahead of normal maintenance cycle). There is usually a delay between emergence of a new phenomenon and its availability as a MeSH descriptor. For example the term “*Maternal Serum Screening Tests*” was introduced in MeSH in 2013, whereas the first articles on the subject were published over 30 years ago [5]. (Of note, a specific term for “cell-free DNA” is currently under consideration for introduction in MeSH.) Therefore, our approach based on text mining is better suited for identifying trends in a timely fashion.

Limitations and perspectives

For text mining purposes, we had to extract text of articles from PDF documents, which are optimized for human readability, rather than automatic text processing. For example, we had to eliminate text of headers and footers to avoid extracting the name of the publisher present on each article as a “frequent term”. Similarly, we had to ignore words containing digits, which resulted in absence of potentially important terms, such as “*b2-microglobulin*”, “*CRISPR/Cas9*”, and many gene names (e.g. “*CHD7*” or “*FGFR3*”). Availability of *Prenatal Diagnosis* corpus in computer-friendly formats, such as XML, would make text mining analyses simpler and more reliable.

As mentioned earlier, we had to manually review terms excluded by our medical term filter and rescue 2.7% of them for analysis, including “*fetal nuchal translucency*” and “*cell-free DNA*”. This is a consequence of limited coverage of Fetal Medicine terms in standard terminologies integrated in UMLS. Recent inclusion of Human Phenotype Ontology [18] into UMLS (version 2015AB) brought some important terms for postnatal phenotypes, but coverage of Fetal Medicine remains limited [8].

Conclusion

Through a bibliometric analysis of articles published in *Prenatal Diagnosis*, using text-mining techniques, we were able to identify trends in Fetal Medicine over the past 10 years. These trends are related to diagnostic techniques (Cytogenetics is progressively replaced by non-invasive techniques based on Genomics) and to study design (Fetal Medicine increasingly relies on scientific methods, including statistics and bioinformatics).

Our bibliometric analysis identified trends that are consistent with those identified by experts (about recent diagnostic techniques), but also identified other interesting trends (about study design), and provided an account for terms exhibiting falling trends and stable terms. In practice, bibliographic analysis and expert opinion are complementary approaches to identifying trends in Fetal Medicine.

PubMed/MEDLINE indexing based on MeSH offers limited granularity and a delay that is not compatible with identification of trends in a rapidly evolving domain, such as Fetal Medicine. We observed that coverage of Fetal Medicine, in MeSH, and standard terminologies integrated in UMLS is limited. List of terms identified through our text mining analysis could be basis for developing a terminology for Fetal Medicine. The list of terms and their evolution profiles are available upto request to the authors.

In summary, a bibliometric analysis based on text mining effectively supports identification of trends over time. This scalable approach is complementary to analyses based on metadata or expert opinion.

Acknowledgements

This work was supported in part by Intramural Research Program of NIH, National Library of Medicine, French Gynecology and Obstetrics Association (*Collège National des Gynécologues et Obstétriciens Français*), and Philippe Foundation.

References

- [1] L.D. Arnold, M. Braganza, R. Salih, and G.A. Colditz, Statistical trends in the Journal of the American Medical Association and implications for training across the continuum of medical education, *PLoS One* **8** (2013), e77301.
- [2] D.W. Bianchi, T. Van Mieghem, L.G. Shaffer et al, In case you missed it: the Prenatal Diagnosis section editors bring you the most significant advances of 2013, *Prenat Diagn* **34** (2014), 1-5.
- [3] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res* **32** (2004), D267-270.
- [4] R. Bolli, The 10 Most Read Articles Published in Circulation Research in 2015, *Circ Res* (2016).
- [5] H.S. Cuckle, N.J. Wald, and R.H. Lindenbaum, Maternal serum alpha-fetoprotein measurement: a screening test for Down syndrome, *Lancet* **1** (1984), 926-929.
- [6] J. Deprest, A. Ghidini, T. Van Mieghem et al, In case you missed it: the Prenatal Diagnosis editors bring you the most significant advances of 2015, *Prenat Diagn* **36** (2016), 3-9.
- [7] B.H. Faas, A. Ghidini, T. Van Mieghem et al, In case you missed it: the Prenatal Diagnosis editors bring you the most significant advances of 2014, *Prenat Diagn* **35** (2015), 29-34.
- [8] M. Girdea, S. Dumitriu, M. Fiume et al, PhenoTips: patient phenotyping software for clinical and research use, *Hum Mutat* **34** (2013), 1057-1065.
- [9] Y. Hong, Q. Yao, Y. Yang et al, Knowledge structure and theme trends analysis on general practitioner research: A Co-word perspective, *BMC Fam Pract* **17** (2016), 10.
- [10] L.J. Jensen, J. Saric, and P. Bork, Literature mining for the biologist: from information retrieval to biological discovery, *Nat Rev Genet* **7** (2006), 119-129.
- [11] F. Li, M. Li, P. Guan et al, Mapping publication trends and identifying hot spots of research on Internet health information seeking behavior: a quantitative and co-word biclustering analysis, *J Med Internet Res* **17** (2015), e81.
- [12] K.J. Lim, D.Y. Yoon, E.J. Yun et al, Characteristics and trends of radiology research: a survey of original articles published in AJR and Radiology between 2001 and 2010, *Radiology* **264** (2012), 796-802.
- [13] R. McLean, K. Mendis, B. Harris et al, Retrospective bibliometric review of rural health research: Australia's contribution and other trends, *Rural Remote Health* **7** (2007), 767.
- [14] M.E. Norton, B. Jacobsson, G.K. Swamy et al, Cell-free DNA analysis for noninvasive examination of trisomy, *N Engl J Med* **372** (2015), 1589-1597.
- [15] S. Perez, V. Laperriere, M. Borderon et al, Evolution of research in health geographics through the International Journal of Health Geographics (2002-2015), *Int J Health Geogr* **15** (2016), 3.
- [16] A. Pinter, Changing Authorship Patterns and Publishing Habits in the European Journal of Pediatric Surgery: A 10-Year Analysis, *Eur J Pediatr Surg* **25** (2015), 353-358.
- [17] R Core Team, R: A Language and Environment for Statistical Computing (version 3.2.2), *R Foundation for Statistical Computing, Vienna, Austria*, (2015).
- [18] P.N. Robinson, S. Kohler, S. Bauer et al, The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease, *Am J Hum Genet* **83** (2008), 610-615.
- [19] L.C. Sayres, M. Allyse, M.E. Norton et al, Cell-free fetal DNA testing: a pilot study of obstetric healthcare provider attitudes toward clinical implementation, *Prenat Diagn* **31** (2011), 1070-1076.
- [20] M. van Wesel, Evaluation by citation: trends in publication Behavior, Evaluation Criteria, and the strive for high impact publications, *Sci Eng Ethics* **22** (2016), 199-225.
- [21] G.T. Venable, B.A. Shepherd, C.M. Loftis et al, Bradford's law: identification of the core journals for neurosurgery and its subspecialties, *J Neurosurg* **124** (2016), 569-579.
- [22] J. Weiner, tagcloud: Tag Clouds. R package version 0.6, (2015).
- [23] H. Wickham, ggplot2: Elegant Graphics for Data Analysis., *Springer-Verlag New York* (2009).
- [24] H. Wickham, stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.0.0, (2015).
- [25] H. Wickham and R. François, dplyr: A Grammar of Data Manipulation. R package version 0.4.3, (2015).
- [26] A. Zietman, The Red Journal's Top Downloads of 2014, *Int J Radiat Oncol Biol Phys* **93** (2015), 4-6.

Address for correspondence

Olivier Bodenreider, MD, PhD

8600 Rockville Pike, 38A/09S904, Bethesda, MD 20894, USA

Phone: (301) 827-4982, E-mail: olivier.bodenreider@nih.gov

The Portal of Medical Data Models: Where Have We Been and Where Are We Going?

Sophia Geßner^a, Philipp Neuhaus^a, Julian Varghese^a, Philipp Bruland^a,
Alexandra Meidt^a, Iñaki Soto-Rey^a, Michael Storck^a, Justin Doods^a, Martin Dugas^a

^a Institute of Medical Informatics, University of Münster, Münster, Germany

Abstract

To address current key problems of medical documentation: lack of transparency, overwhelming amount of medical contents to be documented and missing interoperability, the Portal of Medical Data Models (<http://medical-data-models.org/>) was established in 2012. Constantly evolving, four years later, the portal displays more than 8900 medical data models with more than 250000 items, of which 84 % have been semantically annotated with UMLS codes to support interoperability. Giving an update on new functions and contents of the portal, two additional export formats have been implemented, allowing the reuse of forms such as HL7's framework Fast Health Interoperability Resources (FHIR) Questionnaires, as well as the OpenDataKit format. Future projects include the implementation of an ODMtoOpenClinica converter, as well as supporting the reuse of forms with Apple's ResearchKit and Android's ResearchStack.

Keywords:

Surveys and Questionnaires, Semantics, Clinical Trial

Introduction

The Portal of Medical Data Models (MDM), established in 2012, is a constantly evolving and fast-growing German and European information infrastructure for medical research and healthcare [1,2]. The multilingual platform allows the upload, download, discussion, ranking and reuse of medical questionnaires or documentation forms. These “medical data models” are created in Operational Data Model (ODM) format, developed by the Clinical Data Interchange Standards Consortium (CDISC) [3]. ODM is XML-based and represents the standard exchange format for research metadata in order to facilitate interoperability amongst various software systems [4]. Apart from ODM, the portal offers various download formats, enabling the import of metadata into different medical information systems. To improve interoperability and data integration, medical concepts are semantically annotated with Concept Unique Identifiers (CUIs) from the Unified Medical Language System (UMLS), developed by the U.S. National Library of Medicine [5]. The identification and coding of medical concepts is manually performed by medical experts respecting the published coding principles [6]. The creation of medical data models has been standardized, allowing uniform semantic annotation, using ODMedit [7]. The integrated editor ODMedit proposes possibly matching data elements, which have been defined and semantically annotated before and may be reused.

The value of interoperability in the United States was evaluated in 2005, showing that the fully standardized exchange and interoperability of health care information between various providers has the potential to save 77.8 billion dollars per year [8].

The German Federal Ministry of Education and Research has launched a funding initiative to establish data integration cen-

ters, for which the portal provides an infrastructure for a standardized exchange of medical data models between participating consortia in order to analyze their medical (meta-) data landscapes [9].

The overwhelming amount of data documented in medicine, as well as the number of distinct medical concepts in a clinical terminology, such as Systematized Nomenclature of Medicine (SNOMED CT) [10], indicate the “astronomical” number of potential medical data models used by health professionals [11]. Up to now, most of the medical forms and data from clinical trials remain unpublished. This leads to a cost and time intensive process of re-developing and re-implementing case report forms (CRFs) for clinical studies and documentation forms in EHR systems. Researchers, practicing physicians and their patients are calling for “open (meta-) data”, starting to consider clinical trial data as a public good [12]. Unpublished data hampers the systematic review and reproduction of published results leading to redundancies in clinical research as well as uncertainties in patients and treating physicians. MDM addresses the lack of transparency by granting open access to all of its contents. Furthermore its contents may be downloaded under different Creative Commons Licenses, allowing sharing and adapting the material for different purposes.

In this paper we will analyze the current contents and describe new functions of the portal. Further research objectives give an overview of the user's activities, planned additional download formats, functions and further research based on MDM. The research questions of this paper can be summarized as followed:

1. What kind of medical contents and functionalities are available in the system and how did it evolve over time?
2. Who is using the system and what export formats are selected most frequently?

Methods

Architecture of the Portal and its contents

The technical background of the portal and the editor have been described before [7,13]. To summarize, the portal has been implemented in Ruby on Rails. The data models are stored on a web server. After uploading an ODM file, its structure is additionally stored in a MySQL database on the server. [7,13]

The medical data models are sustainably archived by the University Library of Muenster [14].

Designing medical data models with ODMedit

As data models represent an “interoperable image” of actual medical forms, they are created in the CDISC ODM format using original templates from clinical trials, registries, routine

documentation forms, common data elements (CDEs), data standards or patient-reported outcomes. Medical experts, such as physicians, medical documentalists and medical students, create and semantically annotate forms with UMLS codes using ODMedit [7]. Currently 14 medical experts with more than 120 hours per week are creating and annotating medical data models manually. The UMLS CUI of a medical concept is stored as an alias property of the items or codelist items. If a medical concept can be described by a single CUI, it is called a pre-coordinated concept. If a medical concept needs more than one CUI to represent its full meaning, the process is called post-coordination. Uploaded medical data models are reviewed and verified for accuracy before being released for public use.

Analysis of the contents

Database queries were performed using the MySQL Workbench 6.3. To determine the time course of items and medical data models offered by the portal, the cumulative number of items and data models from November 2011 until November 2016 were analyzed. Furthermore, the UMLS codes mapped to the items and the Medical Subject Headings (MeSH) descriptors of the category C for diseases mapped as keywords to the medical data models were analyzed to display the medical contents of the portal.

To display the portal's user distribution, a world map was configured using Leaflet, OpenStreetMap and Mapbox. The retrieval of geolocations of the users' most recently used IP addresses was conducted using freegeoip.net. To analyze the geolocation of potential users who didn't register but visited the portal and accessed forms, Apache Log-Files of the previous eight months were assessed, ignoring Internet bots.

An online user survey was conducted in February 2016 using LimeSurvey. The online questionnaire, available in English and German, contained questions on missing functions and further export formats requested by users.

The new functions of the portal were reviewed: the table of contents was expanded and modified after reviewing and analyzing the contents of the portal. By choosing a subentry of the table of contents, the search function is called with predefined terms, similar to a manual search. Two additional export formats were implemented.

Results

Contents of the Portal of Medical Data Models

As of November 2016 the Portal of Medical Data Models contains 8948 active forms, 16794 forms in total, with a total of 256751 items. The analysis of the time course of medical data models shows that the number of medical data models as well as the number of items are continuously growing (see Figure 1). Since December 2015, about 400 new medical data models per month were uploaded to the portal. As for the number of items, approximately 10000 are added per month. 84 % of the items are semantically annotated with UMLS codes, representing their medical concepts. The amount of semantically annotated items totals to 216586 items. The three most common UMLS codes tagged to items, representing medical contents, are "C0022885" for "Laboratory Procedures", "C0031809" for "Physical Examination" and "C0006826" for "Malignant Neoplasms".

Other UMLS codes that were assigned most frequently are mainly administrative concepts, such as "C0011008" – "Date in time" or "C2348585" – "Clinical Trial Subject Identifier".

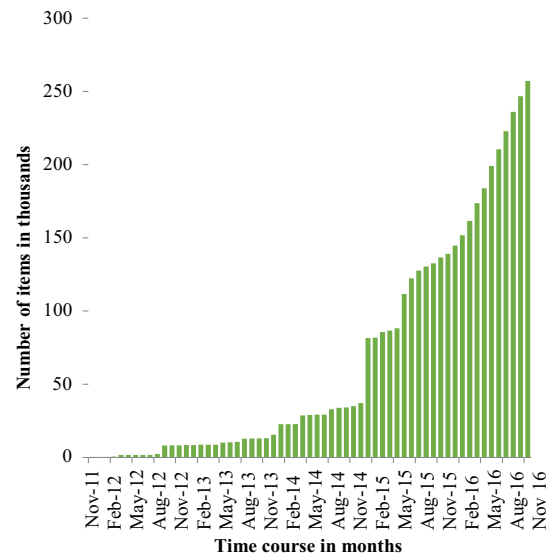


Figure 1 - Time course of items in active medical data models represented in the Portal of Medical Data Models

The results of the analysis of MeSH descriptors is presented in a logarithmic scale in figure 2. A total of 13747 assigned descriptors were identified. The prevailing majority of medical data models is related to oncology. More than 4300 medical data models have been tagged with a MeSH descriptor of the category C04 for neoplasms. The second most used MeSH terms are branches of the category C20 for Immune System Diseases, including diseases such as multiple sclerosis and asthma. The third most represented category is C17 for Skin and Connective Tissue Diseases.

The portal offers contents in multiple languages. Though most 7731 forms (86 %) are monolingual, either in English or German, 1199 forms (13 %) are available in two languages, mainly in English and German. A total of 14 forms (0,16 %) are available in more than two languages, ranging from 3 languages up to 30 languages. The most common language is English with 8414 forms, followed by German with 1858 forms, French with 13 forms and Spanish, Italian and Polish with 5 forms.

Online Survey and User's distribution

The analysis of the users requests for further export formats revealed the demand for HL7's framework Fast Health Interoperability Resources (FHIR) Questionnaires, combining the most advantageous features of HL7 v2, v3 and CDA for exchanging health data [15]. Other export requests were OpenClinica, an electronic data capture system, using Excel templates to import metadata and OpenDataKit (ODK), a tool-set for mobile data collection.

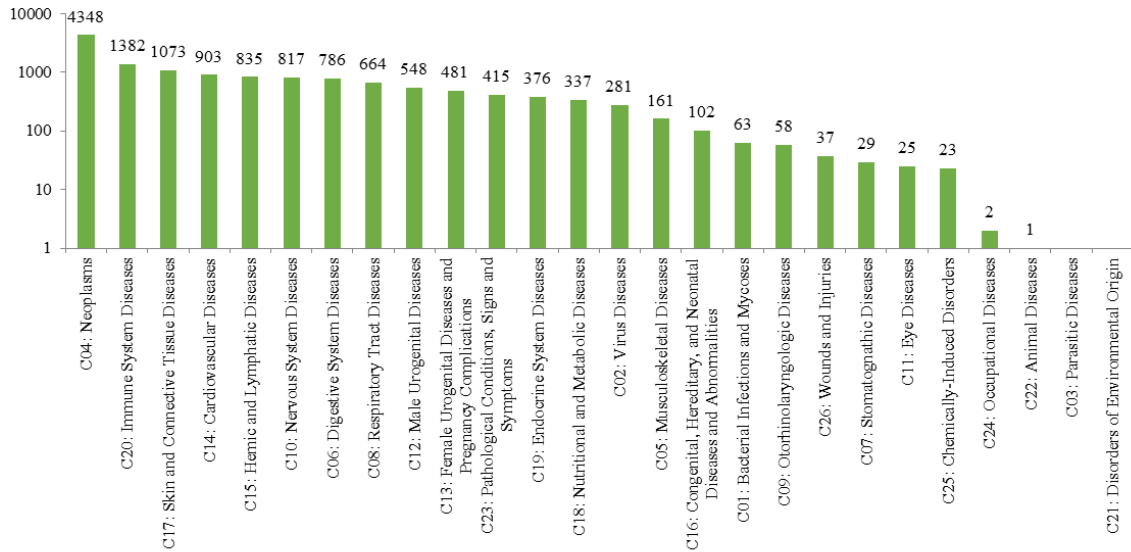


Figure 2 - MeSH descriptors of the category C for diseases assigned to active forms in a logarithmic scale

Currently the Portal has 611 registered and active users, distributed worldwide (Figure 3, green pins). Analyses of users visiting the portal without registration showed the distribution as presented in Figure 3 by the blue pins. The majority of users are located in Central Europe and the United States.



Figure 3 - Geographical distribution of MDM Portal users worldwide. The users' locations are focused on central Europe and North America. The green pins represent registered users ($n=611$), the blue pins are unregistered users ($n=5307$).

The total numbers of users, registered and potential are displayed in Table 1, grouped by their geographical origin.

Table 1 - Total number potential users' page views, grouped by their geographical origin

Region	Number of registered users	Number of form visits by unregistered users
Europe	454	3186
Northern/Central America	102	1514
South America	4	60
Asia	34	424
Africa	6	65
Oceania	11	58

Further functions and export formats

In response to the users' requests, two new export formats were implemented. A converter, transforming forms from ODM to FHIR Questionnaire Resources was implemented in Java and integrated into the portal. This included identifying equivalent elements and mapping these ODM elements to the elements of the FHIR Questionnaire Resources [16]. A similar approach was conducted implementing the requested converter ODM to ODK.

Table 2 - Number of downloads for each export format since February 2016

Format	Number of downloads	%
ODM	254	35.2%
PDF	99	13.7%
REDCAP	86	11.9%
CSV	50	6.9%
PDF WITH COMMENTS	48	6.7%
FHIR-XML	37	5.1%
FHIR-JSON	32	4.4%
SQL	26	3.6%
CDA	23	3.2%
XLSX	23	3.2%
MACRO	13	1.8%
SPSS	13	1.8%
ADL	9	1.2%
R	8	1.1%

Table 2 gives an overview of the most frequently downloaded formats. Since the implementation of the ODMtoFHIR converter in February 2016, the FHIR questionnaire format is at 5% for FHIR-XML and 4% for FHIR-JSON as one of the six most common download formats. Most forms were downloaded in ODM (35%), PDF (14%) and REDCap (12%). The least downloaded formats were ADL [17] and R with 1%.

The table of contents (see Figure 4) was modified and adapted to the medical contents of the portal. It contains 7 main categories, arranging the contents by type of documentation within the medical data model, such as "Clinical Trial", "Routine Documentation" or "Patient-Reported Outcomes". Furthermore the contents are indexed by specialty under

“Specialty-related forms”, containing 16 subitems, such as “Internal Medicine”, “Neurology” and “Surgery”.

Table of contents:	Expand all
1 Clinical Trials	
2 Routine Documentation	
2 Routine Documentation	
2.1 Patients Admission	
2.2 Medical History	
2.3 Physical Examination	
2.4 Assessment Classifications/Scores	
2.5 Nursing Report	
2.6 Apparative Diagnostics	
2.7 Laboratory	
2.8 Pathology/Histology	
2.9 Patient's Information/Consent forms	
2.10 Treatment/Therapy	
2.11 Operative Report	
2.12 Progress	
2.13 Discharge Summary/Letter	
2.14 Aftercare	
3 Registries	
4 Quality Assurance	
5 Data Standards and Common Data Elements	
6 Patient-Reported Outcome	
7 Specialty-Related Forms	

Figure 4 - Table of contents of the Portal of Medical Data Models. The subentry “Routine Documentation” is expanded, showing the contents of the medical documentation in clinical routine. The subitems are chronologically arranged, following the patients path through the clinic from admission to discharge.

Discussion

The analysis of the contents of the Portal of Medical Data Models show major progress over the past year.

In November 2015 the portal contained about 4300 models with about 136500 items. One year later, the number of medical data models more than doubled to the amount of over 8900, increasing by about 400 data models per month. The amount of items also almost doubled to 256751 items, reaching an upload rate of about 10000 items per month. Our goal is to reach an upload rate of 600 medical data models per month in order to ensure the availability of data models in a timely manner. Furthermore the portal shall cover most contents in medical research and clinical practice. Difficulties concerning the open access to metadata in routine patient care and clinical and epidemiological research are still far from being resolved. Small steps in the right direction are being made, as journals like the New England Journal of Medicine have committed to data sharing [18].

To enable semantic interoperability of medical data models, they are semantically annotated. With an annotation coverage of 84%, the portal provides the possibility to access more than 246500 semantically annotated items for reuse. The most frequently assigned UMLS codes with medical content, “laboratory procedures” and “physical examination” represent very well two of the most common medical items documented and examined in clinical trials. It has to be mentioned that semantic annotation is not a trivial task. From our experience, manual review by medical experts is highly required. Even respecting coding principles, mapped UMLS codes still differ between independent coders. When generating CDEs, this still represents an issue, leading to the inevitable, time-consuming

process of “code-cleaning”. Steps towards uniform semantic annotation are being conducted using the integrated editor ODMEdit [7]. Furthermore, the impact of coding principles and ODMEdit on inter-coder reliability are being evaluated.

With regards to the representation of disease entities in the portal, we are able to show that a wide range of disease entities is already represented. The majority of contents are related to oncology, explaining the third most frequent semantic code assigned (“Malignant Neoplasms”). This is in line with the contents of ClinicalTrials.gov, a registry for clinical studies, maintained by the U.S. National Institutes of Health [19]. It currently lists a total of more than 231900 studies, 25 % of which are associated with “Neoplasms”. The European EU Clinical Trials Register currently displays more than 29300 studies [20]. In accordance with the worldwide registry ClinicalTrials.gov, the amount of clinical trials related to oncology represents, with more than 7000 studies, about 24% of the contents. A great amount of medical data models in the portal have been created based on the trial inclusion and exclusion criteria, indicated in the study record details of studies on ClinicalTrials.gov. With a focus on oncologic, neurologic and cardiovascular diseases, this resembles the numerical distribution of contents. So far, the portal does not contain medical data models, mapped with a MeSH descriptor of the tree branches C21 for disorders of environmental origin and C03 for parasitic diseases. Only one medical data model from the tree branch C22 for animal disease was found.

Another medical research repository is the National Institute of Health’s (NIH) Common Data Element Repository, maintained by the U.S. National Library of Medicine. It currently offers 2175 case report forms in four export formats (ODM, two NIH/CDE schemata and SDC). The main research domain focus is on neurological disorders (NINDS) with 1409 elements, as well as on PhenX measures, a consented amount of 627 “standard measures of phenotypes and exposures for use in research” [21]. This is only a small subset of medical forms, compared to the amount of forms offered by our portal. Nevertheless, one must note, that the contents of the NIH CDE Repository consist mainly of previously consented CDEs. Considering the costly and time-consuming task, the creation of CDEs poses, involving the close cooperation of various research communities, the NIH CDE Repository contains already a substantial number of reusable forms. By presenting an image of the current state of documentation in various medical backgrounds, the Portal of Medical Data Models may contribute to a faster, more efficient and effective way to create CDEs, serving as the infrastructure used to identify and generate CDEs. In course of their doctoral thesis, six medical students are currently doing research on various disease entities. To identify CDEs used in myeloid leukemia, the portal already presents a solid and feasible foundation [22].

By adapting the table of contents, clinicians may be able to get a quick overview over the topics of their specialty. To approach the needs of users, we are constantly implementing further export possibilities. As the user survey showed, there is a great need for an export of medical forms to the Open-Clinica metadata import format. The converter is currently being developed and will be available in the shortterm.

Furthermore we soon will offer an export to Apple’s ResearchKit, in order to support the reuse of our contents in studies conducted by mobile devices via mobile applications. We are planning the implementation of an export “ODMtoResearchStack” as it represents the equivalent to Apple’s ResearchKit for surveys conducted by Android users.

A problem with the current implementation in Ruby on Rails is the lack of scalability and easiness of maintenance. To face these problems the system is currently being re-developed in Java EE, which is planned to go online in the course of 2017.

Once the contents of the portal represent most of the currently used medical data models, the transnational creation of data standards and CDEs may be accelerated and will support the interoperability of clinical data, leading to harmonized documentation, improving cross study comparisons and meta-analyses.

Conclusion

Transparency, interoperability and huge amounts of data and metadata are crucial issues in medical research, approached by the Portal of Medical Data Models. This paper gives an overview of the new contents of the portal, such as the representation of an increasing amount of medical data models and items as well as new functions such as converters from ODM to FHIR or ODK. Additionally, the presentation of content is now structured according to the origin and specialty that the medical data models are related to. The user survey revealed the demand for further export features, which will be developed in the near future.

Acknowledgements

This work is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant DU 352/11-1). We would like to thank all students who contributed to this project, especially Jan Kenneweg.

References

- [1] MERIL - Mapping of the European Research Infrastructure Landscape, http://portal.meril.eu/converis-esf/publicweb/research_infrastructure/3574, accessed 26 October 2016.
- [2] Deutsche Forschungsgemeinschaft (DFG), RIsources - The Research Infrastructure Portal, http://risources.dfg.de/detail/RI_00396_en.html, accessed 22 November 2016.
- [3] Clinical Data Interchange Standards Consortium, Operational Data Model (ODM) - XML, <https://www.cdisc.org/standards/foundational/odm>, accessed 14 October 2016.
- [4] V. Huser, C. Sastry, M. Breymaier, A. Idriss and J.J. Cimino, Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM), *Journal of Biomedical Informatics* **57** (2015) 88–99.
- [5] U.S. National Library of Medicine, Unified Medical Language System (UMLS), <https://www.nlm.nih.gov/research/umls/>, accessed 21 October 2016.
- [6] J. Varghese and M. Dugas, Frequency analysis of medical concepts in clinical trials and their coverage in MeSH and SNOMED-CT, *Methods of Information in Medicine* **54** (2015) 83–92.
- [7] M. Dugas, A. Meidt, P. Neuhaus, M. Storck and J. Varghese, ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository, *BMC Medical Research Methodology* **16** (2016) 65.
- [8] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, D.W. Bates and B. Middleton, The value of health care information exchange and interoperability, *Health affairs* (2005) W5-10-W5-18.
- [9] Federal Ministry of Education and Research, Medical Informatics Funding Scheme - Networking data – improving health care, http://www.gesundheitsforschung-bmbf.de/_media/Flyer_Medizininformatik_englisch_barrierefrei.pdf, accessed 21 October 2016.
- [10] U.S. National Library of Medicine, SNOMED CT, <https://www.nlm.nih.gov/healthit/snomedct/>, accessed 25 October 2016.
- [11] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland and J. Varghese, Portal of medical data models: information infrastructure for medical research and healthcare, *Database: The Journal of Biological Databases and Curation* **2016** (2016).
- [12] Institute of Medicine (IOM), Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington, D.C: The National Academies Press (2015).
- [13] B. Breil, J. Kenneweg, F. Fritz, P. Bruland, J. Doods, B. Trinczek and M. Dugas, Multilingual Medical Data Models in ODM Format: A Novel Form-based Approach to Semantic Interoperability between Routine Healthcare and Clinical Research, *Applied Clinical Informatics* **3** (2012) 276–289.
- [14] Universitäts- und Landesbibliothek Münster, miami ulb Münster: Medical-Data-Models.Org, <https://miami.uni-muenster.de/Record/4f9faa6d-cbad-4c96-aade-2306f76bb642>, accessed 27 October 2016.
- [15] FHIR v1.0.2, <https://www.hl7.org/fhir/summary.html>, accessed 25 November 2016.
- [16] J. Doods, P. Neuhaus and M. Dugas, Converting ODM Metadata to FHIR Questionnaire Resources, *Studies in Health Technology and Informatics* **228** (2016) 456–460.
- [17] P. Bruland and M. Dugas, Transformations between CDISC ODM and openEHR Archetypes, *Studies in Health Technology and Informatics* (2014) 1225.
- [18] J.M. Drazen, Data Sharing and the Journal, *N Engl J Med* **374** (2016) e24.
- [19] U.S. National Institutes of Health - ClinicalTrials.gov, <https://clinicaltrials.gov/>, accessed 12 December 2016.
- [20] European Medicines Agency - Clinical Trials Register, <https://www.clinicaltrialsregister.eu/ctr-search/search>, accessed 12 December 2016.
- [21] PhenX Toolkit, <https://www.phenxtoolkit.org/>, accessed 30 November 2016.
- [22] J. Varghese, C. Holz, P. Neuhaus, M. Bernardi, A. Boehm, A. Ganser, S. Gore, M. Heaney, A. Hochhaus, W.-K. Hofmann, U. Krug, C. Muller-Tidow, A. Smith,
- [23] Weltermann, T. de Witte, R. Hehlmann and M. Dugas, Key Data Elements in Myeloid Leukemia, *Studies in Health Technology and Informatics* **228** (2016) 282–286.

Address for correspondence

Sophia Gefner
Email: [Sophia.Gessner\(at\)uni-muenster.de](mailto:Sophia.Gessner(at)uni-muenster.de)

Perceiving the Usefulness of the National Cancer Institute Metathesaurus for Enriching NCIt with Topological Patterns

Zhe He^a, Yan Chen^b, James Geller^c

^a School of Information, Florida State University, Tallahassee, FL, USA

^b Department of Computer Information Systems, BMCC, City University of New York, New York, USA

^c Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, USA

Abstract

The National Cancer Institute Thesaurus (NCIt), developed and maintained by the National Cancer Institute, is an important reference terminology in the cancer domain. As a controlled terminology needs to continuously incorporate new concepts to enrich its conceptual content, automated and semi-automated methods for identifying potential new concepts are in high demand. We have previously developed a topological-pattern-based method for identifying new concepts in a controlled terminology to enrich another terminology, using the UMLS Metathesaurus. In this work, we utilize this method with the National Cancer Institute Metathesaurus to identify new concepts for NCIt. While previous work was only oriented towards identifying candidate import concepts for human review, we are now also adding an algorithmic method to evaluate candidate concepts and reject a well defined group of them.

Keywords:

Biomedical Ontologies; Vocabulary; Quality Assurance

Introduction

Biomedical ontologies, terminologies and controlled vocabularies are widely used in a variety of healthcare information systems for encoding healthcare data such as diagnoses, laboratory test results, patient-reported problem lists, and billing statements. They are also used to facilitate knowledge management, data integration, decision support, and biomedical natural language processing. The development and curation of biomedical ontologies are mostly driven by ontology engineers and subject matter experts. As domain completeness is one of the key properties of a biomedical ontology [1], new concepts need to be included as they are needed by the users. As manual curation is costly and time consuming, automated or semi-automated methods for identifying new concepts that are relevant to the domain of an ontology are therefore in high demand.

The National Cancer Institute (NCI) Thesaurus (NCIt) is an important reference terminology in the cancer domain. It currently contains over 100,000 concepts that are hierarchically organized into 19 distinct hierarchies relevant to cancer research, such as *neoplastic diseases* and *molecular abnormalities*. NCIt is a central reference terminology of NCI's Enterprise Vocabulary Services (EVS) [2]. The EVS leverages both an internal quality assurance (QA) team and external participation in the development and QA of NCIt. Outside contributors can suggest new concepts or terms for NCIt, which will be reviewed, validated, and incorporated into

it, based on NCI's content development and editing guidelines.

In previous work [3; 4], we have developed a topological-pattern-based method to demonstrate the vertical density differences across pairs of source terminologies in the Unified Medical Language System (UMLS), the most comprehensive biomedical terminological system in existence, developed by the U.S. National Library of Medicine [5]. Leveraging the topological patterns in the UMLS, we have identified potential new concepts for SNOMED CT [3; 4]. Figure 1 illustrates the simplest case of a topological pattern, called a $k:1$ trapezoid where $k=2$. In this case, both Terminology 1 (T1) and Terminology 2 (T2) contain Concept A and Concept B. In other words, Concept A and Concept B in T1 and T2 have the same UMLS Concept Unique Identifier (CUI). There is one intermediate concept X between Concept A and Concept B in T1 but no intermediate concept in T2. Upward pointing arrows indicate IS-A links.

One can argue that Concept X may be missing in T2. However, the intermediate concept(s) from T1 may not be needed in T2 according to human judgment. If $k > 2$, there is more than one intermediate concept in T1. In this paper, T2 is always NCIt. The final decision whether the intermediate concepts should be included in NCIt or not always has to be made by its curators.

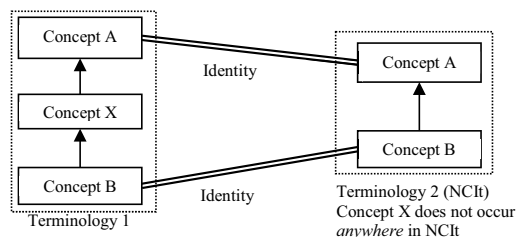


Figure 1 – A hypothetical 2:1 trapezoid between Terminology 1 and Terminology 2 (NCIt)

Recently, we applied this method to identify concepts in the UMLS that could enrich NCIt [6]. The National Cancer Institute Metathesaurus (NCIm) is a terminological system with the same structure as the UMLS, but with more cancer related content [7]. Therefore, a natural question arises – “Is NCIm a better source for utilizing topological patterns to identify new concepts for NCIt than the UMLS Metathesaurus?” Because NCIm and UMLS have much overlap, we were especially interested in concepts suggested for import by topological patterns that exist only in the UMLS

but not in the NCI_m and vice versa. The latter case would indicate a new source for candidate concepts compared to previous research.

One difficulty with our previous approach is that all evaluations of candidate concepts have to be performed by a human expert. In this paper, we were investigating ideas for algorithmically rejecting some of the candidates, reducing the work load for the human expert. We formulated the following hypotheses about reasons for missing intermediate concepts in $k:1$ trapezoids in NCI_t:

- The parent concept (A in Figure 1) has only one child (B in Figure 1) in NCI_t. Therefore, an intermediate concept is not needed to organize the conceptual content.
- The intermediate concept is a synonym of the parent concept (A in Figure 1) or the child (B in Figure 1) in NCI_t.
- The curators of T1 and NCI_t made different modeling decisions.

The contributions of this work are twofold: 1) We compare the usefulness of the NCI_m and UMLS Metathesaurus as data sources for using topological-pattern-based methods to identify new concepts for NCI_t; 2) We propose a practical method to algorithmically reject some concepts for import, thereby reducing the work load of the human expert.

Background

UMLS Metathesaurus (Meta)

The UMLS Metathesaurus (Meta), which integrates over 190 biomedical terminologies, is the most comprehensive biomedical terminological system in existence [5]. It maps over 12 million terms into over 3.2 million concepts.

NCI Metathesaurus (NCI_m)

The NCI Metathesaurus (NCI_m) is a wide-ranging biomedical terminological system that covers most terminologies used by NCI for clinical care, translational and basic research, public information and administrative activities [7]. It maps about four million terms from more than 75 source vocabularies into two million concepts that represent their meaning. Importantly, it has a monthly update cycle, ensuring the timely update of its core terminology, NCI_t. It covers most of the public domain terminologies of the NLM's UMLS Metathesaurus as well as many other biomedical terminologies created by or of interest to NCI and its partners such as RadLex. Some vocabularies such as SNOMED CT were also shrunk in the NCI_m.

Methods

Identifying Candidate Terminologies

In this work, we compared the effectiveness of using NCI_m versus UMLS Meta to identify potentially useful concepts for enriching the conceptual content of NCI_t. We used the August 2015 version of the NCI_m and the 2015AA version of the UMLS Meta. Both NCI_m and UMLS 2015AA contain SNOMED CT US (March 31, 2015 version). The NCI_t version in NCI_m is 2015 08E, whereas the NCI_t version in the UMLS is 2014 03E. The main criteria for selecting a candidate terminology for this research include: 1) the

terminology must be in English; 2) the terminology must be organized with an IS-A hierarchy backbone; 3) the terminology must exhibit sufficient overlap in content with NCI_t; and 4) the terminology must exist in both UMLS Meta and NCI_m. We first identified seven English source terminologies with "PAR" (parent-child) relationships and "INVERSE IS_A" relationship attributes, including SNOMED CT, Foundational Model of Anatomy Ontology (FMA), Universal Medical Device Nomenclature System (UMD), RadLex (a radiology lexicon), University of Washington Digital Anatomist (UWDA), MGED Ontology (MGED), and Gene Ontology (GO). Out of these seven terminologies, RadLex and MGED are not in the UMLS Meta. UWDA is part of FMA and was therefore excluded.

Identifying and Analyzing $K:1$ Trapezoids

We first identified all the $k:1$ trapezoids in NCI_m and UMLS. The NCI_m, which is based on the UMLS, may contain cycles of IS-A relationships [8]. We eliminated the cycles in the trapezoid by detecting repeating Concept Unique Identifier (CUIs) in the IS-A paths.

After we identified all the $k:1$ trapezoids in the NCI_m and the UMLS, we calculated the number of trapezoids for each kind and the number of intermediate concepts in T1. Note that multiple parents may lead to overlapping trapezoids with the same intermediate concept. We eliminated duplicate intermediate concepts in the results.

Manual Review of the 2:1 Trapezoid Samples

To compare the utility of the topological patterns in the NCI_m and the UMLS Meta for identifying new concepts for enriching NCI_t, we took a random sample of 50 2:1 trapezoids between SNOMED CT and NCI_t that can be found in the UMLS but not in the NCI_m (Sample 1), as well as a random sample of 50 2:1 trapezoids between SNOMED CT and NCI_t that can be found in the NCI_m but not in the UMLS (Sample 2). We combined the two samples and randomized the order. The terminology expert (YC) investigated the content of both SNOMED CT and NCI_t using the Neighborhood Auditing Tool (NAT) for the UMLS [9], and assessed whether the intermediate concepts in SNOMED CT should be suggested for inclusion in NCI_t or not. The terminology expert chose one of the following three options: 1) the intermediate concept (Concept X in Figure 1) in T1 should be imported into NCI_t; 2) the intermediate concept should not be imported into NCI_t; and 3) the intermediate concept *may* be imported to NCI_t. For the options 2) and 3), the terminology expert was also asked to give rationales for making such a choice. The NAT tool (<http://nat.njit.edu/>) allows an auditor to concentrate on a single focus concept and its neighborhood (i.e., parents, children, siblings), thereby well meeting the need of this study. We will report the manual review results and the reasons for options 2) and 3) in the Results section.

Identifying More Complex Topological Patterns

There could also be one or more intermediate concepts in T2 in a topological pattern. We therefore defined $M:N$ trapezoids as topological patterns in which both T1 and T2 have both Concept A and Concept B, but there are $M-1$ intermediate concepts between A and B in T1 and $N-1$ intermediate concepts between A and B in T2. The intermediate concepts in T1 do not appear anywhere in T2 and vice versa. $M:N$ trapezoids are a generalization of $k:1$ trapezoids. The relationships among intermediate concepts in $M:N$ can be categorized into the following three types: 1) an intermediate concept in T1 can be a parent/child of an intermediate concept

in T2; 2) an intermediate concept in T1 can be a synonym of an intermediate concept in T2; 3) T1 and T2 have alternative classifications, indicating two different ways of conceptualizing a domain that are both valid but not immediately compatible [3]. A trapezoid may also indicate an error in one of the two terminologies. In a recent publication [10], we provided an estimate of the difficulty faced by a domain expert in a concept import task. In this paper, we merely identified the M:N trapezoids ($M \geq 2$, $N \geq 2$) in both the NCIm and the UMLS. The analysis of the relationships among intermediate concepts in M:N trapezoids is beyond the scope of this work.

Results

2:1 Trapezoids in the NCIm and the UMLS Meta

Table 1 shows the number of 2:1 trapezoids identified in the NCIm and the UMLS Meta. As shown in Table 1, notably fewer 2:1 trapezoids between SNOMED CT and NCIt were identified in NCIm than in the UMLS. In these 2:1 trapezoids, 1,019 distinct intermediate concepts can be found in both NCIm and the UMLS; 890 distinct intermediate concepts were found in the UMLS but not in the NCIm; and 174 intermediate concepts in the NCIm were not in the UMLS. This may be due to the fact that the NCIm contains a newer version of NCIt than the UMLS Meta. Only a small number of trapezoids could be found between terminologies other than SNOMED CT and NCIt. In the subsequent analysis, we will focus on the trapezoids between SNOMED CT and NCIt.

Table 1 – Number of 2:1 Trapezoids Identified in NCIm and UMLS Meta

Candidate Terminology	NCIm		UMLS Meta	
	# of Trapezoids	# of Intermediate Concepts	# of 2:1 Trapezoids	# of Intermediate Concepts
SNOMED CT	2,308	1,193	3,894	1,909
FMA	115	55	112	55
GO	57	38	54	37
UMD	2	2	1	1

Figure 2 shows the histogram of semantic types of 890 intermediate SNOMED CT concepts that were identified in the 2:1 trapezoids in the UMLS Meta but not in the NCIm. Semantic types with fewer than 10 concepts are not shown in the figure.

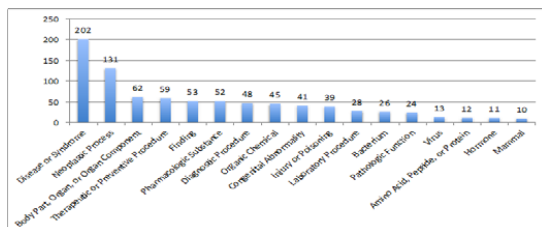


Figure 2 – The Semantic Types of 890 Intermediate SNOMED CT Concepts in the 2:1 Trapezoids in the UMLS, but not in the NCIm.

Figure 3 shows the histogram of semantic types of 174 intermediate SNOMED CT concepts that were identified in

the 2:1 trapezoids in the NCIm but not in the UMLS. All the semantic types except for ENZYME also appear in Figure 2.

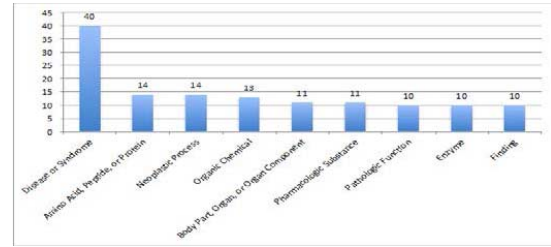


Figure 3 – The Semantic Types of 174 Intermediate SNOMED CT Concepts in the 2:1 Trapezoids in the NCIm, but not in the UMLS.

K:1 Trapezoids in NCIm and the UMLS Meta

Table 2 shows the number of trapezoids and the number of unique intermediate concepts identified in k:1 ($k > 2$) trapezoids between SNOMED CT and NCIt in both NCIm and the UMLS Meta. As the value of k increases, the difference between NCIm and UMLS becomes larger, indicating that NCIm has significantly downsized SNOMED CT and removed concepts that are unnecessarily granular for the cancer domain.

Table 2 – Number of Trapezoids and Unique Intermediate Concepts from SNOMED CT Identified in k:1 Trapezoids ($k > 2$) in NCIm and the UMLS

Kind	NCIm		UMLS Meta	
	# of Trapezoids	# of unique intermediate concepts	# of Trapezoids	# of unique intermediate concepts
3:1	967	876	2,386	1945
4:1	356	532	1,686	1495
5:1	256	462	1,509	1143
6:1	209	373	1,357	1001
7:1	82	172	1,032	687
8:1	15	44	700	424
9:1	3	20	461	261

Manual Review of the Samples of 2:1 Trapezoids

According to the review results of YC, in Sample 1 (2:1 trapezoids found in the UMLS but not in the NCIm), 20 intermediate concepts (40%) should be imported into NCIt, 27 intermediate concepts (54%) should not be imported into NCIt, and 3 intermediate concepts (6%) may be imported into NCIt. In Sample 2 (2:1 trapezoids found in the NCIm but not in the UMLS), 20 intermediate concepts (40%) should be imported into NCIt, 23 intermediate concepts (46%) should not be imported into NCIt, and 7 intermediate concepts (14%) may be imported into NCIt. For the cases in which the intermediate concepts should not be imported, the major reasons for rejecting the import are: 1) The term of the intermediate concept is a synonym of the parent/child concept in NCIt; 2) the parent concept has a single child in NCIt; it was felt that creating a structure of one child with one grandchild contradicts the idea of hierarchically organizing concepts in an ontology into groups of similar concepts; and 3) NCIt and SNOMED CT are using two different categorizations. For example, SNOMED CT models the concepts by sites but NCIt does not.

For the cases of possible import, the rationale is that SNOMED CT is more granular than NCIt, i.e., the parent concept has more children/descendants in SNOMED CT than in NCIt. We list the review results of the mixed sample in Table 3. Even though both samples have the same percentage of trapezoids that can contribute concepts to NCIt, Sample 2 has a higher percentage of intermediate concepts that *may* be imported into NCIt than Sample 1 (14% vs. 6%).

Table 3 – Results of Sample Review by the Terminology Expert

Class	Reason	Sample 1 (%)	Sample 2 (%)
Should be imported	--	20 (40%)	20 (40%)
Should not be imported	1) Synonyms	5 (10%)	4 (8%)
	2) Single child	8 (16%)	12 (24%)
	3) Different categorizations	14 (28%)	7 (14%)
May be imported	--	3 (6%)	7 (14%)

Figure 4 illustrates a case of reason 1). In this case, the term of the intermediate concept *Blood Cell Count* is a synonym of the concept *Complete Blood Count* (C0009555) in NCIt. Therefore, there is no need to add the intermediate concept in SNOMED CT to NCIt.

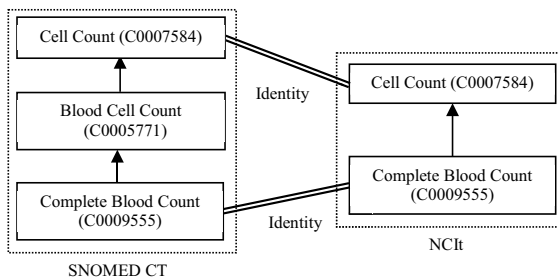


Figure 4 – A 2:1 trapezoid between SNOMED CT and NCIt in the UMLS in which the term of the intermediate concept in SNOMED CT is a synonym of the child concept in NCIt.

Figure 5 illustrates a case of reason 2) in which the parent concept *Retinitis* (C0035333) has only one child - *Cytomegaloviral Retinitis* (C0206178) in NCIt. It is therefore not recommended to add an intermediate concept in NCIt.

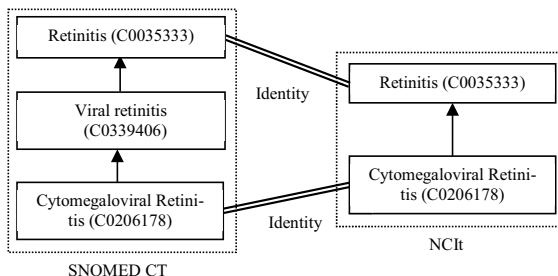


Figure 5 – A 2:1 trapezoid between SNOMED CT and NCIt in the NCIm, in which the parent concept *Retinitis* has only one child in NCIt.

Figure 6 illustrates a case of reason 3) for rejecting a concept import. In this 2:1 trapezoid, the intermediate concept in

SNOMED CT is modeled by sites but NCIt does not follow the same design. Thus, the intermediate concept should not be imported.

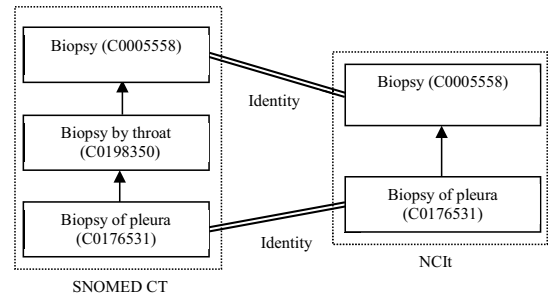


Figure 6 – A 2:1 trapezoid between SNOMED CT and NCIt in the NCIm in which the intermediate concept is modeled by site.

Figure 7 illustrates a 2:1 trapezoid in which the intermediate concept *Digestive System Disorder* in SNOMED CT should be imported into NCIt.

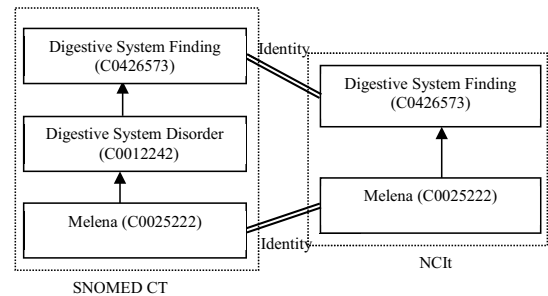


Figure 7 – A 2:1 trapezoid between SNOMED CT and NCIt in the NCIm in which the intermediate concept should be imported into NCIt.

Automatic Rejection of Import from K:1 Trapezoids

As discussed in the manual review section, one of the major reasons why an intermediate concept in a 2:1 trapezoid should not be imported into NCIt is that the parent concept in the trapezoid has only one child in NCIt. To assess the scale of this problem, we investigated all the $k:1$ trapezoids between SNOMED CT and NCIt in both the UMLS and NCIm to find the percentage of parent concepts in the $k:1$ trapezoids with only one child in NCIt. As can be seen in Table 4, 8.7% of parent concepts in 2:1 trapezoids in the UMLS have only one child in NCIt, whereas 7.5% of parent concepts in 2:1 trapezoids in the NCIm have only one child in NCIt. As k increases, the percentage drops fast. About 5% of intermediate concepts identified in 2:1 trapezoids with such a structure can be automatically rejected for import. Another reason for rejecting an import is that the term of the intermediate concept is a synonym of the parent/child concept in NCIt. However, it is not easy to fully automate the detection of such cases because certain judgments of synonymy need to be made by human experts.

M:N Trapezoids in NCIm and the UMLS Meta

Table 5 shows the number of $M:N$ trapezoids ($M, N \geq 2$) between SNOMED CT and NCIt in the NCIm and the UMLS, respectively. Due to the space limit, we only show the

numbers of 2:N and 3:N trapezoids. A smaller number of M:N trapezoids was found in the NCIm than in the UMLS. Whether a higher percentage of trapezoids is useful for concept enrichment or not warrants further investigation.

Table 4 – Automatic Rejection of Concept Import When the Parent Concept has a Single Child in NCIt

Kind	# of Parent Concepts in UMLS with Single Child in NCIt	# of intermediate concepts that can be rejected	# of Parent Concepts in NCIm with Single Child in NCIt	# of intermediate concepts that can be rejected
2:1	84 (8.7%)	89 (4.7%)	55 (7.5%)	57 (4.8%)
3:1	24 (5.6%)	50 (2.6%)	13 (5.2%)	30 (3.4%)
4:1	6 (3.1%)	21 (1.4%)	2 (0.2%)	6 (1.1%)
5:1	0 (0%)	0 (0%)	1 (0.2%)	8 (1.7%)
6:1	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 5 – Number of M:N trapezoids ($M, N \geq 2$) identified between SNOMED CT and NCIt in NCIm and the UMLS

Kind	# of Trapezoids NCIm	# of Trapezoids UMLS	Kind	# of Trapezoids NCIm	# of Trapezoids UMLS
2:2	688	1097	3:2	492	841
2:3	296	464	3:3	333	825
2:4	77	170	3:4	170	364
2:5	61	65	3:5	82	189
2:6	32	30	3:6	44	98
2:7	3	3	3:7	12	28
			3:8	1	5
			3:9	2	2
			3:10	1	1

Discussion and Conclusions

In this work, we compared the usefulness of the NCIm and the UMLS Metathesaurus for finding new concepts for NCIt using topological patterns. We found that even when limiting ourselves to 2:1 trapezoids the NCIm can possibly contribute 1,193 concepts, on the same order of magnitude as the UMLS Metathesaurus, 1,909. We further developed a method to automatically reject $k:1$ trapezoids for concept imports in which the parent concept has only one child in NCIt. For the UMLS Metathesaurus 4.7% intermediate concepts can be automatically rejected, while for the NCIm it is 4.8% based on 2:1 trapezoids alone.

The automatic rejection scenario may be justified in the following manner. When looking at an ontology from the bottom up, as opposed to a top down view, the fundamental idea of the IS-A hierarchy in an ontology is to organize the most specific concepts into groups so that group members are more closely related to each other than to concepts from different groups. Thus, the ontology functions as a way to organize concepts in a way that reflects the real world. This process is then repeated at higher levels, so that groups are themselves grouped together into larger groups of concepts for which the same rule holds, group members are more similar to each other than to members of other groups. While nature might force us to accept groups of size “1” at the bottom level of the ontology, there is no reason to create artificial groups containing only one concept that are themselves alone in their groups.

For the $k:1$ trapezoid cases where the intermediate concepts were deemed to be possible imports into NCIt, the final decision will depend on the NCIt curators at two levels. First the curators will have to decide whether this topic area in NCIt is sparse on purpose, or whether it is only sparse because of lack of time and budget. Then the curators will have to decide about every concept individually whether it is desirable for import into NCIt.

A few limitations need to be noted. The version of NCIt in NCIm (2015 08E) is different from the one in the UMLS. Because the NCIt version in the UMLS always falls behind the NCIt version in the NCIm, we were not able to find NCIm and UMLS Metathesaurus releases with the same version of NCIt and SNOMED CT. Nevertheless, both the 2015AA release of the UMLS and the August 2015 version of NCIt contain the SNOMED CT US March 31, 2015 version. In future work, we plan to develop a more robust method that leverages more sophisticated topological patterns to recommend new concepts for NCIt and reject inappropriate ones.

Acknowledgements

This work was partially supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA190779. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

References

- [1] J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, *Methods Inf Med* 37 (1998), 394-403.
- [2] S. de Coronado, L.W. Wright, G. Frago, M.W. Haber, E.A. Hahn-Dantona, F.W. Hartel, S.L. Quan, T. Safran, N. Thomas, and L. Whiteman, The NCI Thesaurus quality assurance life cycle, *J Biomed Inform* 42 (2009), 530-539.
- [3] Z. He, J. Geller, and Y. Chen, A comparative analysis of the density of the SNOMED CT conceptual content for semantic harmonization, *Artif Intell Med* 64 (2015), 29-40.
- [4] Z. He, J. Geller, and G. Elhanan, Categorizing the Relationships between Structurally Congruent Concepts from Pairs of Terminologies for Semantic Harmonization, *AMIA Jt Summits Transl Sci Proc* 2014 (2014), 48-53.
- [5] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res* 32 (2004), D267-270.
- [6] Z. He, Y. Chen, S. de Coronado, et al., Topological-Pattern-based Recommendation of UMLS Concepts for National Cancer Institute Thesaurus, *AMIA Annu Symp Proc* 2016 (2016), 618-627.
- [7] NCI, Homepage of NCI Metathesaurus, in. <https://ncim.nci.nih.gov/ncimbrowser/>
- [8] O. Bodenreider, Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention, *Proc AMIA Symp* (2001), 57-61.
- [9] C.P. Morrey, J. Geller, et al., The Neighborhood Auditing Tool: a hybrid interface for auditing the UMLS, *J Biomed Inform* 42 (2009), 468-489.
- [10] Z. He and J. Geller, Preliminary Analysis of Difficulty of Important Pattern-Based Concepts into the National Cancer Institute Thesaurus, *Stud Health Technol Inform* 288 (2016), 389-393.

Address for correspondence:

Zhe He, PhD.
School of Information, Florida State University
142 Collegiate Loop
Tallahassee FL, 32306

Email: zhe.he@cci.fsu.edu

Shiny FHIR: An Integrated Framework Leveraging Shiny R and HL7 FHIR to Empower Standards-Based Clinical Data Applications

Na Hong, Naresh Prodduturi, Chen Wang, Guoqian Jiang

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

In this study, we describe our efforts in building a clinical statistics and analysis application platform using an emerging clinical data standard, HL7 FHIR, and an open source web application framework, Shiny. We designed two primary workflows that integrate a series of R packages to enable both patient-centered and cohort-based interactive analyses. We leveraged Shiny with R to develop interactive interfaces on FHIR-based data and used ovarian cancer study datasets as a use case to implement a prototype. Specifically, we implemented patient index, patient-centered data report and analysis, and cohort analysis. The evaluation of our study was performed by testing the adaptability of the framework on two public FHIR servers. We identify common research requirements and current outstanding issues, and discuss future enhancement work of the current studies. Overall, our study demonstrated that it is feasible to use Shiny for implementing interactive analysis on FHIR-based standardized clinical data.

Keywords:

Data Interpretation, Statistical; Health Level Seven; Reference Standards

Introduction

Great attention has been paid recently on the need for interoperability in healthcare information systems and healthcare solutions to facilitate the meaningful use of healthcare data. However, current healthcare application practices have been greatly limited by data heterogeneity and data model diversity. Component reuse, service sharing, and data exchange have become common challenges in the development of healthcare applications.

On the one hand, data standardization plays a very important role in clinical and translational data analysis systems, especially when large-scale data are accumulated. On the other hand, applications for clinical statistics and analysis could advance healthcare by connecting clinical data with analytic tools that can support data analysis, and engage clinicians in the process of analysis. This study is motivated by the need for open solutions to support clinical statistics and analysis with scalable standards-based data model.

Over the last decade, a number of organizations have been targeting the development of clinical data standards to improve medical research, such as Health Level Seven (HL7), the Clinical Data Interchange Standards Consortium (CDISC), the Observational Health Data Sciences and Informatics (OHDSI) and the Patient-Centered Outcomes Research Institute (PCORI). Meanwhile, a number of data models have

been developed to support the needs of clinical data standardization [1, 2]. Some of the data models are:

1. The Operational Data Model (ODM), developed by CDISC, is a vendor-neutral, platform-independent format for exchanging and archiving clinical study data [3]. As a specification, ODM supports interchange of data, metadata or updates of both between clinical systems.
2. The HL7 Clinical Document Architecture (CDA) is a document markup standard that specifies the structure and semantics of “clinical documents” for the purpose of exchange between healthcare providers and patients. The most popular use of CDA is for inter-enterprise information exchange, such as that envisioned for a US Health Information Exchange (HIE) [4].
3. The Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM) specifies a data model that enables the transformation of data contained within different observational databases into a common format (data model) as well as a common representation (terminologies, vocabularies, coding schemes). It also enables systematic data analysis and characterization using a library of standard analytic routines that are written based on the common format. The purpose of the CDM is to standardize the format and content of the observational data, so standardized applications, tools and methods can be applied to them [5].
4. The PCORnet CDM is based on many CDMs of other large national research consortia. PCORnet is an innovative initiative of the PCORI. Each PCORnet partner network maps and transforms data to the PCORnet CDM using the same consistent format (i.e., the same variable names, attributes, and other metadata) [6].
5. The Fast Healthcare Interoperability Resources (FHIR) is an emerging clinical data standard developed at HL7 [7], which enables the representation and exchange of the electronic health records (EHR) data in a standard structure. Others clinical related data standard, such as HL7 Reference Information Model (RIM) [8] and OpenEHR [9] also define common data models that provide a way of organizing data into a standard structure.

Comparing the usability and clinical concepts coverage, we chose FHIR as our practicing data model in this study as FHIR has strong executability based on a RESTful service architecture and multiple flexible data exchange formats. As an emerging executable HL7 standard, the implementations of FHIR have already attracted a lot of interests from healthcare

practitioners, such as SMART on FHIR [10], characterized by its simplicity for implementation and its rigorous mechanism for exchanging data.

R is a functional language and environment for statistical and graphical exploration of data sets, which is widely utilized in clinical and biomedical statistics and research. Shiny is a web application framework with a simplified web deployment mechanism requiring no HTML or JavaScript knowledge [11]. Shiny also includes the ability to integrate R functions for graphical and interactive analysis.

Therefore, with the goal of building reusable and extensible clinical statistics and analysis applications, we aim to design, implement and evaluate an integrated framework called Shiny FHIR using the HL7 FHIR standard and the R-powered web application Shiny.

Materials

In total, four major materials are used in this study, comprising FHIR resources for the data modeling, Shiny and R packages for clinical data analysis functions implementation, and ovarian cancer clinical data for the use case implementation.

FHIR Resources

The FHIR specification defines a set of resources and an infrastructure for handling those resources [12]. The core FHIR resources represent a wide range of healthcare related concepts, both clinical and administrative. The extension FHIR resources represent customized clinical concepts. By aggregating (“bundling”) granular clinical concepts, the resources can support a complex clinical data report and data exchange.

The base FHIR Specification is a platform specification—a specification on which a variety of different solutions are built. The FHIR Specification has 3 versions in its developing timeline, which are the Draft Standard for Trial Use (DSTU1), DSTU2 and the final non-draft version STU3. We currently use the officially released version DSTU2.

HAPI FHIR

HAPI FHIR is an open-source implementation of the FHIR specification in Java [13]. It is based on the easy-to-execute principle, but applied to the Java implementation. HAPI FHIR APIs are designed based on the JAXB and JAX-WS APIs. HAPI FHIR defines model classes for every resource type and datatype defined by the FHIR specification. FhirContext acts as a factory for most other parts of the API as well as a runtime cache of information that HAPI needs to operate. Under the FhirContext, HAPI FHIR APIs support the data model operations with the FHIR resources, and the RESTful client interacts with the FHIR server.

Shiny & R

Shiny contains two components: a user-interface script (ui.R) and a server script (server.R). The ui.R script controls the layout and appearance of the Shiny app. The server.R script contains the instructions of the programming logic about how to build the Shiny app.

Ovarian Cancer Clinical Data

Ovarian cancer (OC) is the most lethal gynecological cancer. Most OC patients undergo radical surgery to maximize reduction of initial tumor volume, and then receive at least six cycles of platinum-based chemo-treatments to eliminate remaining tumors. Although most patients are initially highly responsive to chemotherapy, the majority of them succumb to

recurrent disease [14]. In a pilot study at Mayo Clinic, we are particularly interested in identifying the potentially prognostic importance of preoperative complete blood count (CBC) parameters and derived inflammation markers such as neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and monocyte-to-lymphocyte ratio (MLR), as well as a tumor marker named CA125 [15, 16]. To enable such studies, we collected the ovarian cancer clinical data from three clinical databases: Ovarian cancer registry database, Lab-test database and Clinical document management (CDM) system.

Methods

In our study, we designed an integrated framework known as Shiny FHIR for enabling standards-based, interactive clinical data applications (Figure 1). The framework comprises three modules: 1) a FHIR server module that represents the clinical data (here, ovarian cancer data) from different sources with unified FHIR specification; 2) a workflow module focusing on patient-centered data analysis and cohort-based data analysis; and 3) an interactive user interface module using the Shiny web framework and the established FHIR server. The overall goal of the integrated framework is to enable a library of Shiny Apps with extensible and reusable tools to support clinical statistics and analysis leveraging standardized data exchange and access services based on FHIR.

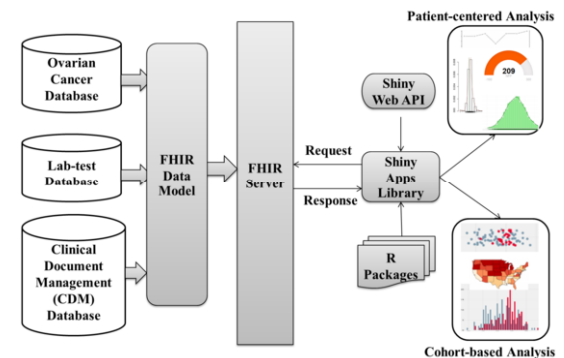


Figure 1 - An Integrated Framework of Interactive Clinical Data Applications

FHIR Data Modeling and Populating

In order to use FHIR for modeling ovarian cancer clinical data to meet the data analysis requirements, we mapped ovarian cancer data to the different types of FHIR resources (Figure 2).

We first extracted demographics, diagnosis, and survival outcomes data from the three databases. 1) Ovarian cancer registry database: Upon diagnosis, the demographic and diagnosis data (e.g. histological types) were manually retrieved, curated and loaded into an ovarian cancer registry database. Clinical outcome data such as survival and recurrence were also routinely updated in the registry. 2) Lab-test database: The CA125 and CBC (e.g., red/white-blood cell counts) test values were retrieved from an institutional lab-test database using the clinical IDs of patients in the OC registry, and returned as time-stamped records with actual test values and additional test-related details. 3) Clinical document management (CDM) system: The chemo-treatment data of

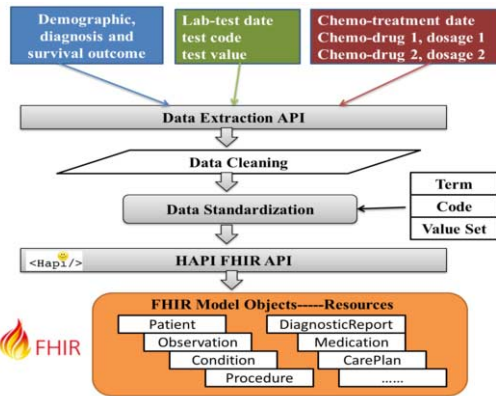


Figure 2 - The Ovarian Cancer Data Processing and FHIR Resources Modeling

identified OC patients were extracted from chemo-treatment documents in an institutional system named CDM, which stores all patients' data regarding clinical visits and the medical services they received. Since the patient data from the different databases are generated during different steps of ovarian cancer diagnosis and treatment, the original data are heterogeneous and describe the patient in different dimensions. In addition, the data formats are also diverse, such as CSV, XML, or combined. However, all the extracted data need to be merged from content to format to support integrative analysis. Since the patient's unique clinic number is the one key guaranteed to connect these objects, we used that number to merge the data into a group of tables in a relational database to support further query and analysis.

Second, we performed basic data cleaning and standardization processes. The data cleaning includes checking and cleaning up redundant data, null-value data, and wrong data generated by data extraction from local databases. In addition, we focused on standardizing the term, code, and value set, and mapping local code to standard code. According to the FHIR resource terminology binding definition, we allocated the Logical Observation Identifiers Names and Codes (LOINC) [17] codes for resource "Observation", and mapped local ovarian cancer related lab test codes manually to LOINC codes. For the chemo-treatment data and diagnosis data, the SNOMED CT is suggested as the binding terminology of "Procedure" and the value set code defined in FHIR is suggested as the code of "Condition". Our methods support assigning of multiple codes for each FHIR resource, therefore, we kept the original codes the Current Procedural Terminology (CPT) [18] for "Procedure", and the International Classification of Diseases (ICD) [19] for "Condition". The purpose of the data preprocessing in this step is to ensure high quality and interoperability of clinical study data.

Third, we analyzed the FHIR data model resources (core resources and extension resources), single resource representations, data elements, data types, terminologies, codes, resource references, search parameters, and data exchange formats (XML and JSON). For example, the FHIR Patient resource structure is shown in Figure 3 [20]. We invoked an open source HAPI FHIR API to populate the ovarian cancer clinical data into the FHIR resources Patient, Observation, Condition, and Procedure.

The next step was to enable secondary use of the ovarian cancer data while simultaneously protecting patient privacy via deidentification. Identifiers were removed from the patient data, and replaced with a set of randomly-generated patient IDs and other privacy objects IDs. We also maintained a mapping table between the local identifiers and FHIR resource identifiers. Thereby, protected health information could be safely utilized for testing and analysis.

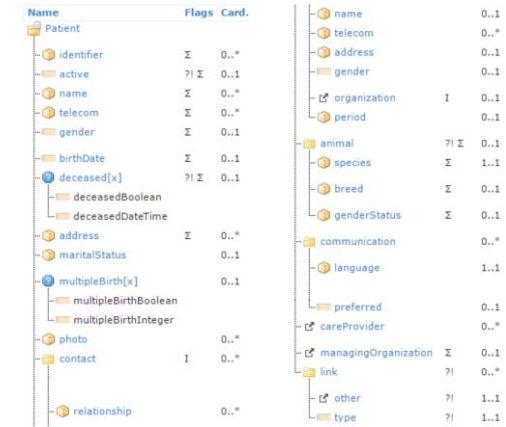


Figure 3 - The Structure Definition of FHIR Resource Patient

Finally, we established a local FHIR server to manage the ovarian cancer clinical data and uploaded the structured FHIR data objects into the FHIR server through HAPI FHIR client API.

The Clinical Data Analysis Workflow Design

Guided by the common clinical data analysis requirements, we designed two general analysis workflows: patient-centered data analysis and cohort-based data analysis (Figure 4).

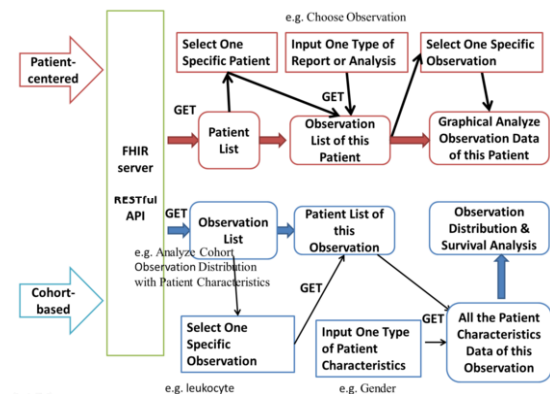


Figure 4 - The Workflow of the Clinical Data Analysis

The patient-centered data analysis workflow provides faceted browsing of all kinds of information oriented to the individual. Patient-specific data derived from different sources are integrated to a single identifier. In the FHIR data model, *Patient* is an independent resource and other resources such as *Observation* and *Procedure* have a property "subject" that is linked to a particular *Patient* object for representing the patient-centered relations.

The cohort-based analysis workflow reflects more common data analysis needs of clinical statistics and studies. In this workflow, the condition/procedure/observation/medication of a cohort is largely measured by the distribution of patient characteristics from different dimensions. The workflow is designed to support a variety of clinical data analysis tasks, including patient observation timeline analysis, patient cohort gender/age distribution statistics, and survival analysis.

The Shiny Implementation on FHIR data

In the prototype implementation step, we adopted the Shiny web framework and developed a prototype platform called Shiny FHIR. The designed workflows and established FHIR server with ovarian cancer data were used to implement the prototype. Shiny FHIR follows a typical structure of Shiny, as shown in Figure 5. The server.R script contains all the workflow functions, and supports the data exchange between Shiny server and FHIR server by using the FHIR RESTful API; the ui.R script controls the layout, appearance and user interactions of the Shiny FHIR in web browser.

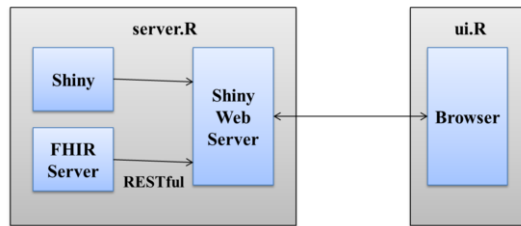


Figure 5 - The System Architecture of Shiny FHIR

From the technical perspective, a series of R packages are aggregated in the workflows to facilitate accessing, parsing, analyzing and visualizing the FHIR data. For example, the 'jsonlite' package supports FHIR JSON structure parsing, the 'dygraph' package is used for the timeline display of observation data values for any particular patient, and survival analysis is supported by the 'survival' package. By drawing on independent contributions from the active R community, the Shiny FHIR platform can facilitate a great number of extensible clinical data analysis applications building on FHIR data. Ideally, as clinical analysis tasks are modularized and the Shiny Apps library matured, a nearly unlimited number of R packages will become available to researchers via FHIR.

Evaluation Methods

We evaluated the feasibility of our framework and prototype by using a number of public FHIR data servers. Our evaluation mainly focused on the adaptability of the framework and the platform to other clinical data following the FHIR standard.

Results

Using the framework described above, we transformed the ovarian cancer clinical data from local EMR into the FHIR-based representations. A total of 61 patient records, 3400 observation rec-ords, 66 condition records and 900 procedure rec-ords and the relations among them were populated into a FHIR server for testing. A total of 4366 "subject" relations were created between the Pa-tient data objects and other data objects (Table 1).

Table 1 - Ovarian Cancer FHIR Resources

FHIR Resources	Instances	Relations
Patient	61	0
Observation	3400	3400
Condition	66	66
Procedure	900	900
Total	4427	4366

For the patient-centered analysis, we designed and implemented a workflow to support the display and discovery of all the clinical data around individual patient, including diagnosis, examination, laboratory test, treatment, etc.

For the cohort-based analysis, we designed and implemented a workflow that measures patient distribution from 11 patient characteristics, comprising age at diagnosis, race, origin of cancer, stage, histology, grade, preoperative CA125 level, ascites at surgery, residual disease, recurrence, and vital status. In addition, the workflow supports a survival analysis task to analyze the prognostic importance of pre-operative lab test values in patients with ovarian cancer. Overall survival rates were estimated via the univariate Cox proportional hazards analysis method. Based on the parameters of observation (for example, red blood cell distribution width), age at diagnosis, deceased time, and the overall survival rates were calculated as the time from the date of diagnosis to the time of death. The prototype and its interactive analysis workflows are displayed as Figures 6 and 7.

Shiny FHIR!

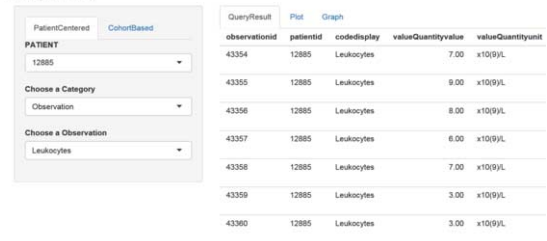


Figure 6 - Patient-centered data analysis

Shiny FHIR!

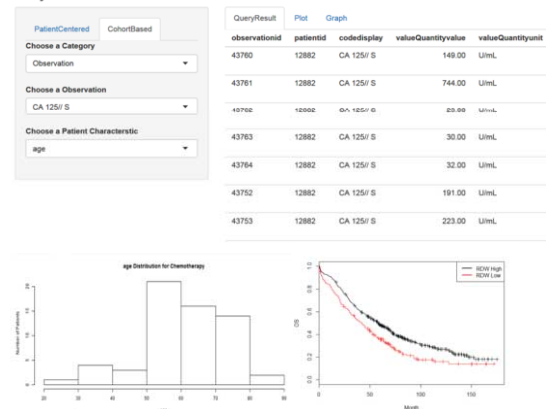


Figure 7 - Cohort-based data analysis

Currently, there are a number of FHIR servers that are publicly available for testing [21]. On the grounds of server stability and data quality, we chose two as testing servers for our system evaluation: the HAPI FHIR server and Michigan Health Information Network Shared Services server. The data published from the two servers were both successfully requested through the FHIR RESTful API. The JSON parsing interface was invoked according to each FHIR resource structure template, and data elements of each type of FHIR resource were successfully parsed and imported into our workflows. The testing demonstrated that our framework and workflow could be generalizable to other clinical data represented in FHIR specification. On the other hand, we encountered some unsatisfying testing results. For example, a number of requests to these servers had issues in terms of the response speed. More critically, data from the test servers contained a FHIR extension representation that was not defined in our FHIR model. Nevertheless, the overall results of the experiment showed that the Shiny FHIR integration approach offers the feasibility of web-based interactive statistical analysis on FHIR-based standardized clinical data.

Discussion

In this study, we developed an integrated framework leveraging Shiny and FHIR, with a prototype implementation and evaluation, aiming to empower standardized clinical statistics and analysis applications. There are two main contributions in this study. First, all the framework and workflow design follows the FHIR data standard, which could be reused in other clinical data domains and enable extensive support for any clinical data that follows the FHIR specification. Second, the data analysis workflow and tools incorporate the experience of clinical statisticians and researchers, and leverage powerful R analytics. Our integrated Shiny FHIR framework could greatly facilitate interactive, user-friendly, and web-accessible data analysis.

One limitation of this research is that the ovarian cancer research use case is currently designed only to test the Shiny FHIR framework, and as such only replicates common clinical research scenarios. Other possible data analysis workflows and customized research discovery scenarios could be performed upon FHIR-specified data, but are not currently supported by our framework.

In addition, there are several technical challenges in our study:

1. When the target data are represented using the FHIR extensions, our framework and APIs need to be enhanced to support the handling of the extensions.
2. Although we developed a records-combination algorithm to overcome the access records limitation problem when requesting FHIR data through the RESTful APIs, the system performance needs to be improved when handling large datasets.
3. Our interfaces are currently developed under FHIR DSTU2 and need to be upgraded to the official FHIR STU3 when it gets finalized and released by HL7.

In the future, we plan to refine the Shiny FHIR Apps and testing datasets in terms of their modularity and reusability, and perform more rigorous evaluations including usability studies for the future development of the Shiny FHIR Apps. We plan to adopt the FHIR STU3 with particular genomic related resource definitions to represent genomic data in our framework. We also plan to build a robust mechanism handling the FHIR extensions. Furthermore, we will continue

to enhance our framework leveraging FHIR terminology service resources to the optimal and meaningful use of standard codes, code systems and value sets.

Conclusion

In this study, we demonstrated that Shiny FHIR is a feasible framework for facilitating interactive analysis on FHIR-based clinical data. We implemented two workflows based on an ovarian cancer analysis use case to enable both patient-centered and cohort-based analyses, and identified common clinical research requirements, outstanding issues, and challenges. The framework can be generalized and applied to other EHR data that follow the FHIR standard. With more and more healthcare data being accessible using the FHIR standard, we believe that our Shiny FHIR framework, with ongoing enhancements and subject to rigorous evaluations, will empower future clinical data applications for statistics and analytics, ultimately advancing clinical and translational discovery.

Acknowledgements

This study is supported in part by NIH grants U01 HG009450 and U01 CA180940. We thank Dr. Paul Kingsbury for his help with language editing.

References

- [1] R.L. Richesson and J. Krischer, Data standards in clinical research: gaps, overlaps, challenges and future directions, *J Am Med Inform Assoc* **14** (2007), 687-696.
- [2] C. Decker, The Past, Present, and Future of Clinical Data Standards, in: SAS Global Forum 2010 Conference, Seattle, WA, USA, 2010, 14.
- [3] Operational Data Model (ODM) - XML, in, 2016.
- [4] The HL7 Version 3 Clinical Document Architecture (CDA®), in, 2016.
- [5] OHDSI Data Standardization, in, 2016.
- [6] PCORnet Common Data Model (CDM), in, 2016.
- [7] FHIR, in, 2016.
- [8] HL7 Reference Information Model, in, 2016.
- [9] OpenEHR Clinical Models Program, in, 2016.
- [10] J.C. Mandel, D.A. Kreda, K.D. Mandl, I.S. Kohane, and R.B. Ramoni, SMART on FHIR: a standards-based, interoperable apps platform for electronic health records, *J Am Med Inform Assoc* **23** (2016), 899-908.
- [11] Shiny-A web application framework for R, in, 2016.
- [12] FHIR: Guide to resources in, 2015.
- [13] HAPI FHIR, in, 2016.
- [14] S.A. Cannistra, Cancer of the ovary, *The New England journal of medicine* **351** (2004), 2519-2529.
- [15] G.J. Rustin, A.E. Nelstrop, P. McClean, M.F. Brady, W.P. McGuire, W.J. Hoskins, H. Mitchell, and H.E. Lambert, Defining response of ovarian carcinoma to initial chemotherapy according to serum CA 125, *J Clin Oncol* **14** (1996), 1545-1551.
- [16] C.C. Sun, D.C. Bodurka, C.B. Weaver, R. Rasu, J.K. Wolf, M.W. Bevers, J.A. Smith, J.T. Wharton, and E.B. Rubenstein, Rankings and symptom assessments of side effects from chemotherapy: insights from experienced patients with ovarian cancer, *Support Care Cancer* **13** (2005), 219-227.
- [17] LOINC, in, 2016.
- [18] CPT Coding Resources, in, 2016.
- [19] WHO-The International Classification of Diseases in, 2016.
- [20] Resource Patient - FHIR Content DSTU2, in, 2016.
- [21] Publicly Available FHIR Servers for testing, in, 2016.

Address for correspondence

Dr. Guoqian Jiang, MD, PhD at jiang.guoqian@mayo.edu.

Key Relation Extraction from Biomedical Publications

Lan Huang^{a,b,d}, Ye Wang^a, Leiguang Gong^{a,c}, Casimir Kulikowski^e, Tian Bai^{a,b,d}

^a College of Computer Science and Technology, Jilin University, Changchun, Jilin, China,

^b Key Lab of Symbolic Computation and Knowledge Engineering Ministry of Education, Jilin University, Changchun, Jilin, China

^c Yantai Intelligent Information Technologies Ltd, Yantai, Shandong, China

^d Zhuhai Laboratory of Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Department of Computer Science and Technology, Zhuhai College of Jilin University, Zhuhai, 519041, China

^e Department of Computer Science, Rutgers, The State University of New Jersey

Abstract

Within the large body of biomedical knowledge, recent findings and discoveries are most often presented as research articles. Their number has been increasing sharply since the turn of the century, presenting ever-growing challenges for search and discovery of knowledge and information related to specific topics of interest, even with the help of advanced online search tools. This is especially true when the goal of a search is to find or discover key relations between important concepts or topic words. We have developed an innovative method for extracting key relations between concepts from abstracts of articles. The method focuses on relations between keywords or topic words in the articles. Early experiments with the method on PubMed publications have shown promising results in searching and discovering keywords and their relationships that are strongly related to the main topic of an article.

Keywords:

Data Mining; Algorithms; PubMed

Introduction

Searching effectively and efficiently for topics within the rapidly growing biomedical literature is an unprecedented challenge for researchers in the field, even with the help of advanced online search tools. The task becomes even more difficult when we try to find relationships between concepts or topics, which represent potentially more interesting and important information than what concepts alone describe. If concept relationships related to the topic can be extracted from the abstract of an article without going through the main body of the article, the results can be organized and presented for quick review based on which more focused further searches can be carried out, with the additional advantage of also increasing the number of articles covered in the search. More importantly, relations extracted from abstracts can be quickly integrated into a knowledge network [1, 2] for ontology-based discovery. Topic mining and relation extraction from unstructured publications such as research papers and case reports has become more feasible and productive over the past decade [3-5]. Most existing methods follow one of two different approaches: pre-specified pattern-based vs more general machine learning.

- Pattern-based relation extraction [6-8] typically relies on syntactically or grammatically pre-specified patterns of sentence components designed to extract one type of relation. To improve specificity and accuracy, generalization of these patterns is needed to

cover most or all expressions likely to be found in a text. These approaches typically suffer from being tightly domain- and even problem- dependent because of customization or over-fitting to the particular characteristics of language used in interpreting a particular class of problems. For different domains, the issues are even more daunting: what works in, for example, financial text retrieval will not be generalizable to biomedical domains, and vice versa.

- Machine learning based relation extraction [9, 10] tackles the topic extraction differently but suffers from related problems of overfitting, in this case, to whatever training set of data is used, which will completely constrain the generalizability of results as in all such learning approaches. If abstracts of all publications in PubMed is of interest, then specifying and obtaining a suitable training set will be an extremely difficult task. Different subfields or subjects in biomedicine usually require models to be trained separately in order to improve performance. However, the cost of this is unavoidably high and strategies of partitioning the datasets are themselves a factor in biasing the results.
- A hybrid method combining the approaches has been developed recently which relaxes the need for a large number of pre-specified patterns of the first approach while reducing the need for large training sets with the goal of achieving results with similar or better accuracies for problems that are general, but expertly circumscribed so as to present a reasonable challenge to how automatic extraction plays off against expert pre-specification.

Most relation extraction methods are developed to extract relations from massive unstructured texts, such as research papers and case reports. Topic mining [11] methods are developed to classify topics for text classification. One main group of topic mining algorithms extract topics from texts, and another group of algorithms labels a body of text with one or more topic words selected from a pre-specified topic list. Topic mining methods discover topic-related words within the text, but do not provide more or deeper information contained in an article.

Relation extraction methods and topic mining methods produce different types of summarizations of an article. The former returns a list of relations, while the latter returns a set of words. Most relation extraction methods are designed to extract all relations in the article, but are not concerned with

whether the relations are linked to the topic of the article. Since these methods process the entire text of each article, the computational load for implementing the algorithms is usually extremely heavy. In addition, it is up to human users to interpret and filter the relevance or significance of the large number of relations returned by these methods, which is usually time consuming and inefficient, since most extracted relations will simply not be related to the main topic of the articles. In contrast, many topic mining methods are reported to achieve impressive accuracies, but fail when the topic of a paper is not clearly described by the set of keywords in the paper. Furthermore, most existing topic mining methods are limited to analyses of single or individual concepts in an article, and do not address important issues of finding relations between key biomedical concepts across all the relevant literature within a large publication repository such as PubMed. To overcome these limitations, we have focused our work on improving key relation extraction using an innovative approach based on a novel combination of topic mining and relation extraction methods. We consider relations of an article connected to its substantive contribution to a topic, as key relations. Key relations are structured, and are composed of a specific relation and the concepts connected to this relation which are the substance of an article.

Our Key Relations Extraction Method (KREM) presented in this paper extracts key topic-related relations from the abstract including the title of an article. KREM processes each article individually and returns key relations in the article that are related to the main topic of the article. A key relation is one that is related to the topic of the article, or is a part of the subject or theme discussed in the article. Key relations together can be viewed as a structured summarization of the article. Our method has been tested using the PubMed repository.

Method

The basic framework of our approach combines both topic mining and relation extraction. For topic mining, a structural statistical model is developed based on which a topic classifier is built. For key relation extraction, we propose and implement a pattern-based algorithm to compute relations between key topic concepts generated by the relation extraction method and by topic mining.

Combining the Methods of Relation Extraction and Topic Mining

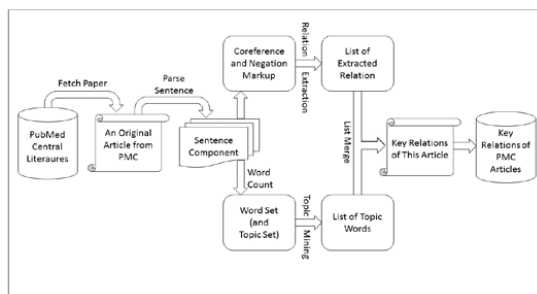


Figure 1—Combined Method of Relation Extraction and Topic Mining

Figure 1 is an overview of the overall framework and workflow of our method using PubMed.

As shown, both topic mining and the initial relation extraction can be carried out in parallel. The relation extraction method computes and produces a bag of relations for an article. In parallel, the topic detection method produces an ordered list of topical words or concepts from the same article, which are ranked by the likelihood of being the topic. The two results are then fused to produce the final result of the key topic relations. Two preprocessing tasks are carried out to produce a list of keywords as topic word candidates.

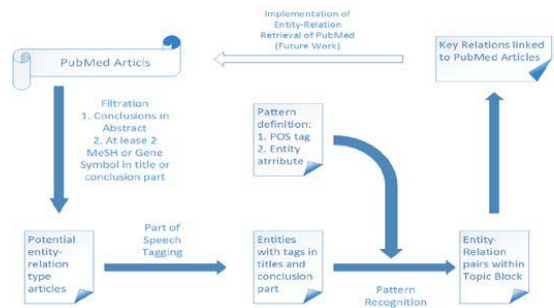


Figure 2—Key Relation Extraction Model

Preprocessing for Keyword Detection Using MeSH and Gene Symbols

We used Medical Subject Headings (MeSH) and Gene Symbols as terms for a standard vocabulary. These two thesauri together include entity names of diseases, chemicals and genes in PubMed articles. We adopted a classic simple method to recognize MeSH and Gene Symbols from articles.

Many MeSH concepts can be recognized from the articles directly with an exception. There is a loose tree structure in MeSH, so some of the parent concepts are also a part of the child concept, for example, “Abdomen” and “Acute Abdomen”. The latter is a child concept of the former, and the entire parent concept is part of the child concept. Both of the concepts can be recognized correctly, however, if the term “Abdomen” were to be considered a false-positive recognition in the sense of being over-general within the more specific child concept of “Acute Abdomen”. We have developed a method to eliminate this kind of error.

We extracted all MeSH concepts and ignored the original tree structure. We compared every pair of concepts in the extracted list, and built a new structured order of MeSH concepts. We labeled words like “Abdomen” as central words and “Acute Abdomen” as an extended word, with other words labelled as isolated words. We thus divided MeSH concepts into 3 groups, with each word in a central word group being connected to its extended words within an extended word group. Therefore, we used members of the center words group and the isolated words group as a thesaurus for recognizing entities in PubMed articles. Once a center word was recognized, we extended the thesaurus to its extended words and if there is a match for one of its extended words, we replaced the center word with its extended word. This way, we increased the level of specificity of what is being recognized to match the level of reference in the article, so as to eliminate

the kind of over-generalization error that would otherwise result.

Gene symbols, on the other hand, are highly specific and unambiguous, and thus entirely different than the anatomical and pathophysiological level concepts in the natural language with English words. With our approach, the mismatch rate is surprisingly low, and we can recognize gene symbols directly from topic blocks in tests, including abstracts.

Handling Abbreviations and Pronouns

Many articles use abbreviations and pronouns instead of MeSH words. These abbreviations and pronouns cannot be recognized as MeSH words, so we replace them with the words they stand for. Abbreviations are explained the first time they are used, we only need to locate the definition of abbreviations and replace them throughout the whole paper with the original MeSH words. As for the pronouns, they do not have a fixed meaning, so we replace them with the nearest noun or noun clause they qualify. Since not every pronoun stands for a MeSH word, we do not need to replace all pronouns. Therefore, we find all MeSH words and use them to replace pronouns related to them as long as the semantic distance between them is short enough.

Structure-Based Probabilistic Model for Topic Concept Recognition

Traditional topic mining or keyword detection methods mostly use word frequency as the main feature. While recognizing the importance of word frequency, we also observe that spatial distribution of words in an article is another very valuable features in text analysis, for topic word detection in particular. The two features are complementary to one another, so we develop a statistical model for topic words and use it in a simple Naïve Bayesian classifier for topic word detection, which appears to be adequate.

The main assumption of this approach is the recognition of the structural partitions in a paper or article, either in the text, such as the Objective-Methods-Results-Conclusion division which is standard for most abstracts found in PubMed, and which we use for developing our statistical model. In doing so, 1000 abstracts were randomly chosen from the PubMed repository. Topic keywords for each abstract were determined from visual examination, with the help of topic mining tools. The frequency of candidate topic keywords appearances in each of the above-mentioned partitions is computed and reported in Table 1.

Table 1-Distribution of Topic Words as Candidate Keywords

	Average Count	Average Appearance rate (%)
Topic Words	1.85	100
Title	1.92	100
Objective/Background	1.92	78.57
Method	1.15	64.29
Result	2.08	85.71
Conclusion	2.23	100

As shown above, the “conclusion” section has the highest frequency of the candidate topic words. The 100% figure shows just how carefully authors made sure to include topic-related candidate keywords in the conclusion section in this particular random sample of abstracts. We believe these

frequency numbers can be taken as a reasonable initial approximation of the structural distributions of topic words to determine the “truly topic-representative” Keywords.

Topic Detection and Relation Prediction Using Naïve Bayesian Classifier

Above, we described how a traditional frequency based method can help discover keywords of an article. The method depends only on the frequency of candidate keywords in the article, and, as shown in Table 1, the location of such candidate keywords in a particular section of structured text can help to identify the topic words which the authors may have intended as keywords. Therefore, we introduced this location of potential topic-keywords within each partition of the text as features and built a Naïve Bayes classifier to discover the most frequently occurring and this likely intended topic-related keywords.

Each keyword identified during the preprocessing phase is extracted, and represented by a feature vector of six specific numerical features, with each feature being a count of occurrence of this keyword in a specific partition.

$$\langle t, ob, m, r, c, e \rangle$$

where t is the title, ob is the objective/background part, m is the method part, r is the result part, c is the conclusion part and e is the number of distinct entities appearing in the same and adjacent sentences for this keyword.

All the potential keywords for an article can thus be represented by a $6 \times N$ matrix with a row for each keyword, forming the feature space for the article, where N is the total number of keywords.

$$\begin{pmatrix} t^1 & ob^1 & m^1 & r^1 & c^1 & e^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ t^N & ob^N & m^N & r^N & c^N & e^N \end{pmatrix}$$

We decided to use Naïve Bayes Classifier for “true” keyword detection for its simplicity and ease in training. Maximum likelihood is used to estimate the parameters of the probabilistic model assuming a normal distribution of samples in the feature space. The prior probabilities of positive and negative sets for a “true” keyword are defined below.

$$P(pos) = \frac{pSize}{pSize + nSize} \quad P(neg) = \frac{nSize}{pSize + nSize}$$

With features assumed to be drawn from a normal distribution, the probability of each feature is $p = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$, where σ is the variance and μ is

the mean value, so we can then define the posterior probabilities for a “true” keyword conventionally as:

Posterior probability of positive entity,

$$posterior(pos) = P(pos)p(title|pos) \times p(abstract1|pos)p(abstract2|pos) \times p(abstract3|pos)p(abstract4|pos) \times p(distinctEntity|pos)$$

Posterior probability of negative entity,

$$posterior(neg) = P(neg)p(title|neg) \times p(abstract1|neg)p(abstract2|neg) \times p(abstract3|neg)p(abstract4|neg) \times p(distinctEntity|neg)$$

When we try to determine the “true” topic keywords of an article, we calculate the feature matrix of this article, and run each row through our Naïve Bayes classifier. As defined previously, each row represents all features of an entity or

potential keyword, with the classifier returning the positive and negative posterior probabilities for each row. We compare the results and look at the posterior likelihood or odds ratio, and if it is greater than 1 we consider the entity to be a “true” topic-relation keyword in the article, and not otherwise. This simplest of Bayesian approaches can be made more sophisticated by using two thresholds for accepting or discarding a potential keyword for an article, with those potential topic words falling in the uncertainty region left for further processing. However, for the present test of our methods, the simple binary accept-reject strategy proved adequate.

Pattern-based Relation Extraction

As mentioned above, we adopted a prespecified pattern-based relation extraction approach as the first phase of our overall hybrid method. We used a sentence parser to divide each sentence into sentence components. Components within a sentence are connected by grammar, and our patterns are built on this. For example, “... pulsatile hemodynamics may be related to amyloidosis and tau-related neurodegeneration...”, is a sentence found in the conclusion of the abstract of PubMed article with PMID 24225352. The words “hemodynamics” and “amyloidosis” are recognized as MeSH entities. We define a pattern $P(E_1, E_2, C_1, C_2)$, where E is the extracted entity, and C is the corresponding sentence component of this entity. In a specific pattern, E represents a single attribute of a MeSH concept or gene symbol. In the example above, this pattern should match the relation, $P(\text{MeSH}, \text{MeSH}, \text{SUBJ}, \text{OBJ})$. For each given pair of components, we have 4 possible combinations of entities. Combined with component pairs, we define 16 base patterns. Other modified patterns then are taken to be variants of these base patterns.

Our pattern-based relation extraction method matches entities within a limited syntactic distance, therefore the entities matched by the method are highly likely to have a relation between them.

Fusion of Topic Mining and Initial Relation Extraction

In general, there are two possible choices of order for running both methods serially. If we run relation the extraction method first, we get a bag of relations. Then, running topic mining second within these relations we would get key relations. However, the words in relations are limited to several trigger words and entities extracted by patterns. This means we run the topic mining method on a small set of words and there is great possibility that the words appearing in relations repeatedly have nothing to do with the topic, because we have no guarantee that the relations are related to topic of the article. If we run topic mining method first, we get an ordered list of words. These words are ordered by the possibility of being the topic of the article. Then we can run the relation extraction method on these words. The results of this are likely to be minimal, as topical entities of an article do not need to function as trigger words. And, if the topical entities of an article cannot trigger relation extraction patterns, we do not get a result. Therefore, running both methods in serial is not ideal.

Experiments and Results

We evaluated our method by testing it against with an implementation of an existing relation extraction algorithm combined with an existing topic mining method called SCRET. Two thousand randomly selected (by PMID) article abstracts are used for the experiment. The sample articles all satisfy the structural assumption mentioned above. Both methods are implemented in Python3 with modules of BioPython [12] and NLTK [13].

SCRET

SCRET produces an empty list of results on 2000 sample articles. Since it is a combination of two algorithms designed for different purposes, this outcome is not a surprise. The method both algorithms adopted dealt with words and phrases differently, and naturally there is little chance that the results of both algorithms can be merged easily. It appears a combination of two algorithms with different purposes is not an effective way to solve such newer, combined problems. However, the idea of combining the two methods is not necessarily irrelevant, we just should not combine the algorithms directly. This is why we developed KREM.

KREM

The 2000 sample articles with conclusions within abstracts all contain over 3 entities within the title and abstract. There are 312 articles that do not have key relations. The rest have key relations and 967 of them have only one key relation, 533 of them have 2 key relations, 137 of them have 3 key relations, 22 of them have 4 key relations, 4 of them have 5 key relations and 5 of them have over 6 key relations. Over 89% of the articles have 1 or 2 key relations. This is consistent with our definition for key relation, if an article is really about an entity-relation, there should be key relations within topic blocks and the number of key relations is limited because the topic of an article is limited. KREM uses topic distribution probabilities in order to avoid the problem in serial implementation in SCRET.

Comparison to RE and SCRET

Traditional pattern-based relation extraction method is not developed for our purposes, but it still can extract relations from PubMed articles. SCRET is a direct combination of relation extraction and topic mining and it mostly satisfies our needs. We compare these two methods to KREM. We selected 50 pieces of articles from PubMed by searching a specific term and manually reviewing each article to make sure these are key relations in each articles. We apply these three methods on this set of articles, the results are shown in Table 2 below.

Table 2– Comparison of relation extraction algorithms

Object	RE	SCRET	KREM
Count of articles with relations	41	0	34
Max count of relations from one article	13	0	4
Average count of relations	4.6	0	1.7
Count of articles with key relations	21	0	21

From Table 2, we can conclude that a direct combination of relation extraction and topic mining cannot provide usable results.

Discussion

We found that lists of results from relation extraction and topic mining differ entirely, and it is difficult to merge them directly, i.e. there are often very few or no similar or identical MeSH words in both lists, so an exact, simple automatic syntactical comparison of the two lists usually produces zero results. Pattern-based relation extraction method finds all relations that fit pre-set patterns, but many of them are not topic-related. Traditional relation extraction methods do not usually recognize key relations well, whereas key relations can be typically found among the results of traditional relation extraction. Based on those relations, we cannot conclude what the main topic of a paper might truly be. KREM returns a limited number of relations, with most of these relations being a true topic or part of the true topic.

Conclusion

We have developed a new practical method for topic word detection and relation extraction to address some of the limitations with existing approaches. It is based on the observation that when a passage of text, such as an abstract of a biomedical article in PubMed, is divided into standard sections, using the statistics of potential topic-related keywords in each separate section can be used as the basis for later Bayesian fusion and then a more reliable second step of topic detection. Preliminary results of testing on a sample of 2000 randomly selected article abstracts show that both methods justify our hypothesis about the structural correlation of topic words, and demonstrate the potential of the new KREM method. Extracted key relations can be viewed as one type of structured summary for an article. Since the semantics of such types of summaries are well understood, they can readily be used for constructing or expanding a knowledge network or ontology. The proposed method has made some important simplifying assumptions about the abstracts of an article being organized in a known structure. This assumption may not apply in longer texts and in some forms of publications, thus limiting the usage of this approach. We are now working on a natural language understanding tool to automatically analyze articles that do not comply with our main assumption of being able to preprocess and partition the article, labeling each partition for further analysis.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61472159, 61572227), Development Project of Jilin Province of China (Nos. 20160204022GX, 20170101006JC, 20170203002GX, 2017C030-1, 2017C033), China Postdoctoral Science Foundation (2014M561293). This work is also supported in part by Premier-Discipline Enhancement Scheme supported by Zhuhai Government and Premier Key-Discipline Enhancement Scheme supported Guangdong Government Funds, and Project 2016192 Supported by Graduate Innovation Fund of Jilin University.

References

- [1] T Bai, L Gong, Y Wang, Y Wang, CA Kulikowski and L Huang, A method for exploring implicit concept related-ness in biomedical knowledge network. *BMC bioinformatics*. 2016 Jul 19;17(9):265.
- [2] Lan Huang, Ye Wang, Yan Wang, and Tian Bai, "Gene-Disease Interaction Retrieval from Multiple Sources: A Network Based Method," *Bio-Med Research International*, vol. 2016, Article ID 3594517, 9 pages, 2016. doi:10.1155/2016/3594517
- [3] A.W. Muzaffar, F. Azam, and U. Qamar, A Relation Extraction Framework for Biomedical Text Using Hybrid Feature Set, *Computational and mathematical methods in medicine* 2015 (2015).
- [4] N.T. Nguyen, M. Miwa, Y. Tsuruoka, T. Chikayama, and S. Tojo, Wide-coverage relation extraction from MEDLINE using deep syntax, *BMC bioinformatics* 16 (2015), 1.
- [5] B. Rink, S. Harabagiu, and K. Roberts, Automatic extraction of relations between medical concepts in clinical texts, *Journal of the American Medical Informatics Association* 18 (2011), 594-600.
- [6] A. Bravo, J. Piñero, N. QueraltRosinach, M. Rautschka, and L.I. Furlong, Extraction of relations between genes and diseases from text and large-scale data analysis: implications for translational research, *BMC bioinformatics* 16 (2015), 1.
- [7] M. Song, W.C. Kim, D. Lee, G.E. Heo, and K.Y. Kang, PKDE4J: Entity and relation extraction for public knowledge discovery, *Journal of biomedical informatics* 57 (2015), 320-332.
- [8] P. Zhang, W. Li, Y. Hou, and D. Song, Developing position structure-based framework for chinese entity relation extraction, *ACM Transactions on Asian Language Information Processing (TALIP)* 10 (2011), 14.
- [9] B. Rink, S. Harabagiu, and K. Roberts, Automatic extraction of relations between medical concepts in clinical texts, *Journal of the American Medical Informatics Association* 18 (2011), 594-600.
- [10] C. Zhang, W. Xu, Z. Ma, S. Gao, Q. Li, and J. Guo, Construction of semantic bootstrapping models for relation extraction, *Knowledge-Based Systems* 83 (2015), 128-137.
- [11] Li Ximing, Ouyang Jihong, Lu You, Topic Modeling for Large Scale Text Data, *Frontiers of Information Technology & Electronic Engineering*, 2015, 16 (6) : 457-465.
- [12] Cock, Peter JA, et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics." *Bioinformatics* 25.11 (2009): 1422-1423.
- [13] Bird, Steven, Edward Loper and Ewan Klein (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.

Address for correspondence

Corresponding Author: Tian Bai, baitian@jlu.edu.cn

Knowledge Management in Health Technology SMEs

Juhamatti Huusko, Sirpa Kuusisto-Niemi, Kaija Saranto

Department of Health and Social Management, University of Eastern Finland, Kuopio, Finland

Abstract

The purpose of this study was to examine knowledge management's (KM) role in small and medium-sized (SMEs) health technology enterprises, which employ fewer than 250 employees. In this study, KM is understood as the ability to achieve competitive advantage by utilizing management knowledge and making it profitable. The health technology enterprises use modern technology to resolve health-related issues. The research data was acquired from Finnish health technology SMEs. The questionnaire was sent to 140 enterprises, generating 25 responses, or a 17.9% response rate. According to the results, health technology enterprises have not adopted KM concepts, nor do they have the necessary resources to do so. SMEs' KM use is informal: information is transferred informally through human interaction, rather than through information systems. In the SMEs, KM is not perceived as important, although it is seen as associated with the enterprise's financial performance through the potential in making the knowledge profitable.

Keywords:

Knowledge Management; Biomedical Technology; Medical Informatics

Introduction

Technological and scientific developments have changed our understanding of business. According to Freedman, Taylorism refers to scientific management focused on the work processes, such as how a certain work phase is efficiently executed. Today, the business environment is changing rapidly, and organizations that apply Taylorism principals are not able to reform quickly enough [1]. The health sector is similarly changing, facing challenges like aging, rising costs, growing demands, demographic changes, and free markets [2]. The health technology industry can offer solutions to these challenges, and could considerably affect health promotion and economic growth. According to the European Union, health technology as an industry consists of diagnostic and treatment methods, medical devices, pharmaceuticals, rehabilitation, prevention methods and supporting systems [3]. The health technology industry can provide new business opportunities and have major economic impacts. For example, in Finland, health technology is an economically significant industry. In 2014, Finnish health technology enterprises exported products and services worth 1.8 billion euros [4].

In general the KM research is in a pre-science phase and is progressing towards academic maturity and normal science [5]. It is argued that KM is affected by economics, sociology, philosophy, and psychology. Three management practices that have significantly influenced KM are information management, quality management, and human factors

movement [6]. Although organizations are not aware of the information's value, KM is a way to satisfy customer needs or even exceed customer expectations [7].

Knowledge management focuses on achieving organizational goals and making knowledge productive, which it does by motivating people and stimulating intrapreneurship, whereby employees act like entrepreneurs. Although SMEs use certain KM instruments to evaluate, acquire, develop, and share knowledge, as well as to determine the knowledge gap, the SMEs do not necessarily call managing information "KM" [8]. Traditionally, KM's research focus has been on larger enterprises [9]. However, both large enterprises and SMEs need KM principles in order to survive in a modern, competitive society. Although the SMEs could achieve greater innovation and productivity by capturing, storing, sharing, and disseminating knowledge, their managers do not see KM investment as sufficiently profitable. The SMEs' KM is usually informal, which means that the knowledge is processed through human interactions rather than information and communications technology (ICT) systems. The managers and owners think that KM is not a business-critical function, and that it is therefore unnecessary for business success. This conflicts with the academic literature, as studies show that both small and large enterprises should invest in KM for the sake of competitiveness maximization, as well as increasing survival probability [10].

Obviously, due to limited resources, SMEs need to be more creative in how they manage knowledge. In fact, SMEs should put effort into managing knowledge because they can use their expertise to gain competitive advantages. The SMEs tend to manage knowledge instantly and without modern ICT. Although they use ICT in various operations, technology's KM use is limited [11]. Knowledge is a primary resource in organizations, and if managed properly, positively affects economic efficiency, innovation, and customer service [12].

Wong and Aspinwall [13] have characterized SMEs' KM as follows:

1. SMEs lack KM understanding, especially regarding its key concepts;
2. SMEs have only slowly adapted formal and systematic KM practices because they do not see it as a high priority.

This study's aim is to examine KM's role in SMEs' health technology sector. The researchers wanted to investigate how these enterprises understand KM, how constantly increasing data is managed, and whether these enterprises use information systems to process this data. To our knowledge, very few studies cover these interests. Although KM in SMEs has gained some attention in academic research, few researchers have been interested in health technology SMEs specifically. To that end, our research question was as follows:

1. What is knowledge management like in the health technology industry?

Health technology industry research is important from both a theoretical and pragmatic perspective. This paper's goal is to increase SMEs' awareness of KM.

Methods

For this descriptive study, a literature search was conducted using databases such as Elsevier and EBSCOhost. The aim was to find scientific papers with themes that engaged with SMEs' KM, specifically within the health technology industry. The following inclusion criteria were applied for study selection:

1. Published in peer-review journals;
2. Focus on SMEs' KM;
3. Published between 1990 and 2016.

The literature search indicated that no measurement tool was created for use in SMEs' KM, specifically within the health technology context. Based on the literature review, a questionnaire was formed using an online survey tool called Webropol. The questions were tightly structured multiple-choice questions and Likert-scale questions. The questionnaire consisted of six demographic questions and nine wider questions concerning themes like informatics and KM. For example, the respondents were asked for their opinions about KM's role in their organization, as well as how information loss is handled in case of retirement. The questionnaire was pretested by nine testers. Some changes were made based on the testing. The survey data was acquired from the Finnish health technology industry SMEs in early 2016. The survey focused on senior SME managers' and board members' views. The online survey was sent primarily to these groups, but snowball sampling was also executed to allow for questionnaire forwarding to other enterprise members. This study's suitable enterprises were reached via their homepages and were included in the survey if they met the inclusion criteria:

1. The enterprises were SMEs (staff headcount < 250, turnover ≤ 50 € m);
2. Their headquarters were located in Finland;
3. They use modern technology to produce services and products that promote health.

The questionnaire was sent to 140 managers, primarily Chief Executive Officers (CEOs) and Chief Technology Officers (CTOs). The respondents received a covering letter where they were informed that their identity or enterprise could not be identified and all the collected data will be destroyed after completing the study. In total, 25 participants responded, generating a 17.9% response rate. The collected data was analyzed using IBM SPSS Statistics 21 software. Due to a relatively low response rate, the analysis used was descriptive statistics. The aim was to provide sample size summaries, such as means and medians, as well as to present demographic information, including age, sex, and education level. The summaries are described using percentages and figures.

Results

In this study, the SMEs were divided into three categories: micro (fewer than 10 employees), small (10–49 employees), and medium (50–250 employees) enterprises. In this study, enterprise size varied between 1–180 employees (Figure 1). A total of 40% of the respondents worked in micro enterprises, while 48% worked in small enterprises, and 12% worked in medium enterprises.

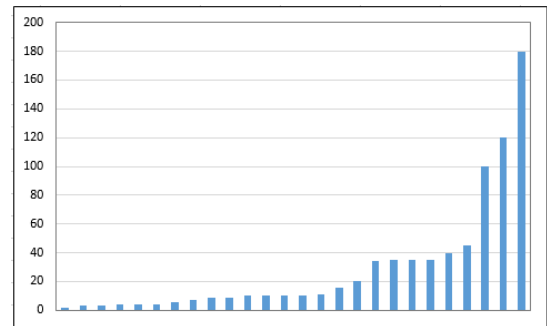


Figure 1 – Enterprise size (number of employees)

The results show that the studied health technology enterprises are male-dominated. Of the respondents, 88% were male and 12% were female. The mean age was 45.8 years, while the median age was 48 years. The youngest respondent was 26 years old, and the oldest was 63 years old. Most of the respondents were CEOs; only two respondents were Chief Information Officers (CIOs), for example. In smaller enterprises, the CEOs had other roles, such as an entrepreneur role, and areas of responsibilities (Figure 2). As the enterprise grows, CEOs can focus on their main role and responsibilities, which are clearly defined. Smaller enterprises do not have a named CIO, so the CEO handles this role. For example, in this study the largest enterprises CEOs main role is to act as a CEO while in the smallest enterprises the CEO may act as a CEO, CIO and as a board member. While the size of the enterprise grows, KM becomes more informal.

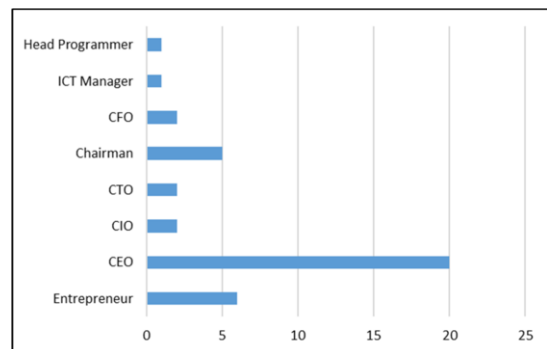


Figure 2 – The respondent's role in the enterprise

The respondents were asked about ICT outsourcing reasons. Approximately 60% of the respondents stated that outsourcing is a way to save costs, 80% stated that outsourcing can help better allocate available resources, and 84% thought that they do not have the necessary knowledge to maintain their ICT infrastructure. Beyond this, 88% of the respondents stated that the ICT information systems provide knowledge associated with competitive advantages and, in general that the ICT provides managers with useful and up-to-date information. The use of modern ICT increases knowledge and information system usability, providing information that informs decision-making. Almost every respondent (96%) highlighted that ICT is significant in knowledge sharing.

In this study, it was essential to examine knowledge's significance in management. To that end, 84% of the respondents noted that their enterprises are dependent on certain employees and entrepreneurs. Therefore, certain employees were seen

as irreplaceable. Meanwhile, 92% of the respondents had an academic degree, which supports the idea that health technology enterprises are knowledge-intensive companies that enhance human capital. In addition, 92% of these respondents stated that the networks, knowledge, and customer relationships were important enterprise resources.

The enterprises were aware of knowledge loss due to employee attrition. In the survey response, 80% of the respondents pointed out that the best way to prevent knowledge loss is to share jobs among multiple employees. Furthermore, apprenticeship was seen as a practical way for reducing knowledge attrition, as apprenticeship is a system whereby experienced and inexperienced employees work side-by-side. Beyond this, a good way to retain the enterprise's knowledge is to cultivate team spirit, which encourages asking for help and the organization culture being open and transparent. A total of 96% of the respondents stated that their enterprise used at least one way to prevent knowledge loss, while one of the respondents stated that this has not been taken into account at all (Figure 3).

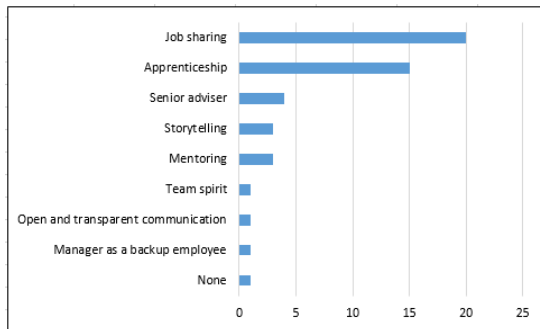


Figure 3 – Ways to prevent knowledge loss due to employee attrition (n = 25)

The respondents were asked for their opinion about KM's role in the enterprise's financial performance. Approximately 96% of the respondents stated that KM is associated with the financial performance, while 84% saw KM as a way to provide management models. Applying KM was seen as a way of providing practical implementation expertise. Furthermore, KM was associated with cost-efficiency, in addition to being linked to value creation and having the potential to be used for identifying strengths, weaknesses, opportunities, and possible threats. Beyond this, respondents stated that KM could be used to make knowledge profitable.

The respondents were asked to describe KM's role in SMEs at a general level (Table 1). The aim was to examine the health technology enterprises managers' views on KM, as well as to compare the findings with those of previous studies. It was found that action is the main way to create, share, transfer, and utilize knowledge in SMEs, with KM being adapted slowly in most of the enterprises participating in the study. The importance of KM received mixed feelings.

Table 1 - Respondents' views of SMEs' KM

	Agree	Undecided	Disagree
SMEs lack KM understanding.	68%	16%	16%
SMEs slowly adapt the formal and systematic KM practices.	72%	16%	12%
SMEs tend to think that KM is not very important.	52 %	20%	28%
The biggest obstacle for KM usage is lack of resources.	64%	8%	28%
SMEs do not utilize technology in KM.	60%	8%	32%
Knowledge is created, shared, transferred and utilized between human interactions.	88%	8%	4%

Discussion

This study's aim was to examine the role of KM in health technology SMEs. The results show that health technology enterprises can be seen as knowledge-intensive enterprises that emphasize human capital. The enterprises' individuals fulfill an essential role, as the enterprises are dependent on certain employees and entrepreneurs. The employees and managers' expertise is relevant for the enterprises, and their jobs require both creativity and problem-solving skills.

In light of this study, health technology SMEs' KM is similar to the previous literature's findings. Health technology enterprises' use of KM does not differ from other SMEs fields. The SMEs have not adopted KM concepts [8–10], and according to these results, the major obstacle to KM's wider use is resource scarcity, which also applies to health technology enterprises.

Thus, SMEs' KM cannot be compared with larger enterprises. Based on the results, it is assumed that SMEs' KM is understood in different ways.

It is important to note that health technology enterprises often offer products and services designed for managing knowledge, and their business models are based on creating, sharing, using, and managing that knowledge. Despite the previous findings, the enterprises themselves have not adopted KM concepts. Based on these results, in some enterprises, product development and marketing absorbs the vast majority of the available resources, which results in scarce resources available for KM.

Beyond this, SMEs' KM tends to be informal, with knowledge being processed in human interactions rather than formally in ICT systems. Although the managers and owners think that KM is essential for their enterprise's success, the resource scarcity forces them to focus on other functions. In this study, health technology SMEs have slowly adapted formal KM practices and do not take full advantage of information technology. Health technology enterprises should invest in KM because it allows the enterprises to gain a competitive advantage [9, 10] and because it is related to financial success and thereby considered an important business function. Applying KM principles can make knowledge productive [7, 11]. It is also essential to recognize that KM can aid in satisfying, or in some cases, even exceed customer expectations [6].

Although the response rate was relatively low, all the respondents completed the questionnaire with no missing data. Therefore, researchers concluded that the questionnaire was well

designed and it functioned properly. The low response rate may originate from the lack of respondents' resources, such as time. It is our assumption that the managers of SMEs are very busy and they do not necessarily have time for extra activities. On the other hand, it is essential to ask, are themes concerning knowledge management somewhat sensitive?

Despite the response rate, this study provides new findings related to the role of KM in SMEs. This study provides insight as to how knowledge is managed in the industry that applies modern technology in a health care setting. It would be necessary, in a further investigation, to carry out surveys and interviews, e.g., mixed methods studies applying quantitative and qualitative data. This method would deepen the understanding of KM in SMEs. In addition, further research should collect data internationally to generate more observations. At the time that our data was collected, Finland had 300 health technology enterprises. In the near future, this number could rise significantly, as the global need grows for high-quality yet affordable health care. Beyond this, technology keeps evolving, and we cannot know for sure what is possible in the future.

In this study, it is worth noting that only 12% of the managers were women. Further studies could examine why women rarely act as managers. Is it because of women's own lack of interests, or perhaps because of general attitudes towards women as managers? It is important to realize that, for example, in Finland, the majority of health care employees are women. It could be necessary to survey more women health technology industry managers and collect their views as well. It is also important to study what kind of efforts are required to manage rapidly increasing rate of new information and how this affects the decision-making in the health technology enterprises. Further studies could also examine how to manage this information overload and how to transfer it into services and products. Examining KM in the health technology sector could enhance our knowledge and help KM progress towards academic maturity [4].

Conclusion

In this study, we examined KM's role in health technology SMEs. We were able to define KM's role in these enterprises. We conclude that these enterprises want to focus on their core business and that due to lack of resources KM is not fully utilized. We believe that it is essential in informatics to examine privately owned health technology enterprises because it is plausible that these enterprises will increasingly process and maintain our health information. As their role in health care grows and as technology evolves, it is essential to require these enterprises to pay more attention to KM processes in order to create, share, use, and properly manage knowledge.

Acknowledgements

We would like to thank all of the health technology enterprises that participated in this study.

References

- [1] D.H. Freedman, Is management still a science?, *Harvard Business Review* **70** (1992), 26–38.
- [2] J. Mantas, Future trends in Health Informatics – Theoretical and practical, In E.J.S. Hovenga and J. Mantas (Eds.) *Global Health Informatics Education* **109** (2004), 114–127.
- [3] European Commission, Health technology assessment, http://ec.europa.eu/health/technology_assessment/policy/index_en.htm, 2016.
- [4] Finnish Health Technology Association (FiHTA), Health technology trade in

2014, http://www.finnishhealthtech.fi/images/stories/industry/reports/terveysteknologian_kauppa.2015.web.pdf, 2015.

- [5] A. Serenko and J. Dumay, Citation classics published in Knowledge Management journals. Part I: articles and their characteristics, *Journal of Knowledge Management* **19** (2015), 401–431.
- [6] L. Prusak, Where did knowledge management come from?, *IBM Systems Journal* **40** (2001), 1002–1007.
- [7] V. Hlupic, A. Pouloudi and G. Rzevski, Towards an integrated approach to knowledge management: 'hard', 'soft' and 'abstract' issues, *Knowledge and Process Management: the Journal of Corporate Transformation* **9** (2002), 90–102.
- [8] R. P. uit Beijerse, Knowledge management in small and medium-sized enterprises: knowledge management for entrepreneurs, *Journal of Knowledge Management* **4** (2000), 162–179.
- [9] S. Durst and Edvardsson, I. R. 2012. Knowledge management in SMEs: a literature review, *Journal of Knowledge Management* **16** (2012), 879–903.
- [10] M.B. Nunes, F. Annansingh, B. Eaglestone and R. Wakefield, Knowledge management issues in knowledge-intensive SMEs, *Journal of Documentation* **62** (2006), 101–119.
- [11] K.C. Desouza and Y. Awazu, Knowledge management at SMEs: five peculiarities, *Journal of Knowledge Management* **10** (2006), 32–43.
- [12] C. Ling-hsing Chang and T.C. Lin, The role of organizational culture in the knowledge management process, *Journal of Knowledge Management* **19** (2015), 433–455.
- [13] K.Y. Wong and E. Aspinwall, Characterizing knowledge management in the small business environment, *Journal of Knowledge Management* **8** (2004), 44–61.

Address for correspondence

Juhamatti Huusko, University of Eastern Finland
Department of Health and Social Management
POB 1627, FIN-70211 Kuopio, Finland
Email: juhamatti.huusko@uef.fi, phone: +358505370525

Effects of Implementing a Tree Model of Diagnosis into a Bayesian Diagnostic Inference System

Satoshi Iwai^a, Yoshimasa Kawazoe^a, Takeshi Imai^a, Kazuhiko Ohe^a

^a Department of Biomedical Informatics, Graduate School of Medicine,
The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Abstract

To estimate a diagnostic probability similarly to experts using answers to interviews, we developed a system that fundamentally behaves as a Bayesian model. For predefined interviews, we defined the sensitivity and specificity related to one or more diagnoses. Additionally, we used a predefined parent-child relation between diagnoses to decrease the number of parameters to set. After calculating the disease probability, we trained the model using the difference of post-test probability between computer calculations and three experts' opinions. We evaluated the effects of setting up tree structures. When using a tree structure, the model trained faster and produced better fitting results than the model without tree structure. Training with multiple raters' training data confused the model. The scores worsened in later epochs. Herein, we present the new method's benefits and characteristics.

Keywords:

Machine Learning; Decision Support Systems, Clinical; Expert Systems

Introduction

Impressive progress has been forthcoming in knowledge representation, machine learning, and data mining for knowledge discovery, and in temporal representation and reasoning [1]. For clinical decision support, some trials have used accumulated data from electronic medical records [2]. However, machine learning methods cannot accommodate rare diseases because appropriate datasets are difficult to obtain. Furthermore, when using EMR data, the possibility exists of learning an incorrect pattern. Machine learning and data mining methods depend heavily on the dataset to be learned. For some aspects of medical knowledge, the introduction of expert knowledge is necessary.

For expert systems that were studied actively in the 1980s, experts crafted knowledge bases [3]. The system was able to accommodate the knowledge implemented on it. However, maintaining knowledge bases for those systems was difficult because medical knowledge has also changed. Therefore, it became necessary to update the knowledge base according to these changes [4]. The most important difference between an expert system and a human doctor is that the expert system is unable to obtain feedback from patients. We devised a system that assumed feedback from manual input by a doctor.

An earlier expert system, INTERNIST-1, used "evoking strength" and "frequency" values that were pre-defined by physicians from reviews of the literature [3]. However, those values included subjective elements that were not based on specific numerical theory. Therefore, the values led to multiple interpretations [5]. Instead of these values, we used

sensitivity and specificity, which have been defined theoretically.

To reach a clinical judgment, physicians usually hypothesize some diagnoses and estimate their respective probabilities. They obtain additional data by conducting interviews. Subsequently, the physicians estimate the diagnoses' respective probabilities. After estimating the probabilities, they choose whether to conduct more tests or to treat the patient without subjecting the patient to risks of further diagnostic tests [6]. This study assesses an approach to simulate specialists' clinical reasoning. We fundamentally used a simple Bayesian model to calculate these probabilities. Additionally, we incorporated the pre-defined tree structure of diseases. This model requires sensitivity and specificity of the interviews to the specific disease, but those values are difficult to determine. We trained these values using the probability of a diagnosis estimated by the experts using answers to patient interviews. We evaluated the effects of implementing a tree model when the model calculated the probability of clinical diagnosis and estimated the sensitivity and specificity.

Methods

Based on a serial Bayesian approach to calculate the respective probabilities of the diagnoses, we developed a computer program to derive differential diagnoses from responses to two-choice interviews. Necessary sensitivity and specificity for the calculations were set in advance for each interview, but these were updated in training from expert opinions. They converge to more appropriate values. To simplify the system and to examine a study of training methodology specifically, we chose dizziness and vertigo as chief complaints. Cases of dizziness and vertigo are often encountered in emergency departments. Reportedly, the history and initial clinical examination of the patients has high sensitivity and specificity for differentiating central and peripheral vertigo [7].

Defining Diagnoses Model

In the emergency department, the granularity of the differential diagnosis differs from that of general outpatients. The differentiation process might be terminated with an upper disease concept. We considered this broader concept of the disease as one differential diagnosis and organized this concept into a tree form (Fig. 1). We used the assumption that the disease concept characteristics are also applicable to child diseases. Statistical characteristics of the relation between interviews and diseases are also inherited to some degree. The model, which was defined from one emergency physician's perspective, was reviewed by another otolaryngologist.

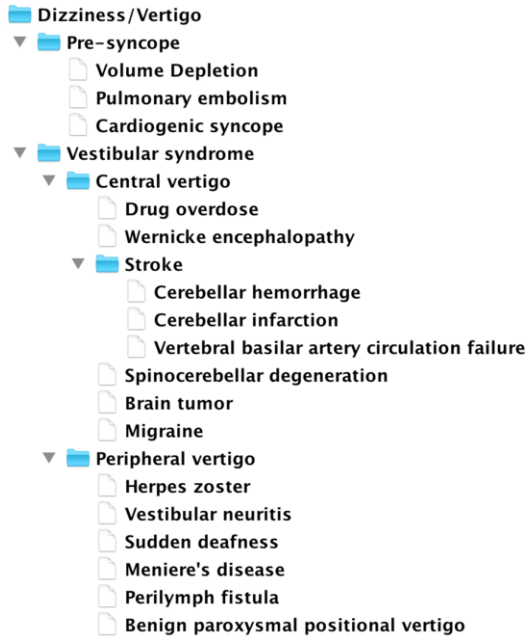


Figure 1 – Tree-form Predefined Diagnoses.

Pre-listing Interviews and Statistical Values

We chose 17 interview queries that are usually posed to dizziness or vertigo patients. Each interview sentence relates to one or more diseases or disease concepts defined above. We designate this relation as a “Diagnosis to Question” (DtoQ) relation. Each DtoQ relation has three attributes: sensitivity, specificity, and an assumed answer. The assumed answer is the answer to a question that is assumed if the disease is present. We regarded the interviews as tests, using them to calculate post-test probability. Table 1 is a 2×2 contingency table showing the frequency of patients who answer the interview. Theoretically, sensitivity is a value that is calculated using formula (1): it is the proportion of those who have a disease, who also respond with the assumed answer to the question. Specificity is a value calculated using formula (2): it is the proportion of those who have no disease, who do not respond with the assumed answer (those who answer *no* if the assumed answer is *yes*). We cannot know these actual values. Therefore, we pre-determined them from a physician’s perspective. Examples of values are presented in Table 2. For this experiment, 27 DtoQ relations were prepared.

Table 1 – 2×2 Contingency Table of Responses

		Disease	
		Present	absent
Interviews	Assumed	a	b
	Not assumed	c	d

$$\text{Sensitivity} = \frac{a}{a+c} \quad (1)$$

$$\text{Specificity} = \frac{d}{b+d} \quad (2)$$

Table 2 – Examples of Pre-listed Interviews and Values

Diagnosis	Interview	Sens	Spec	Ans
Stroke	Do you have a headache or neck pain?	0.6	0.8	1
Migraine	Do you have a headache or neck pain?	0.9	0.4	1
BPPV	Does dizziness improve if you keep still?	0.9	0.6	1
Vestibular neuritis	Has dizziness continued for several days?	0.8	0.4	1
Stroke	Has dizziness continued for several days?	0.7	0.4	1

Sens- predefined sensitivity; Spec - predefined specificity; Ans - Assumed answer; BPPV- Benign paroxysmal positional vertigo

Calculating Certainty

General

We defined the post-test probability as a certainty value. This value was calculated using a serial Bayesian approach, which uses an assumption that each related interview correlates independently to a diagnosis probability. Post-test odds for a diagnosis are calculated by multiplying the pre-test odds. All positive likelihood for interviews for which answers are positive (patient’s answer is the assumed answer) and all negative likelihood for negative ones (opposite answer of the assumed answer) (3). Odds are calculated using formula (4). The positive likelihood ratio and negative likelihood ratio are calculated using formulas (5) and (6). Finally, the certainty value is calculated using formula (7). We cannot estimate the disease probability precisely. Therefore, we assumed pretest probabilities of 0.1 for all diseases.

$$\begin{aligned} \text{post test Odds} &= \prod LR^+(\text{if test} = \text{Pos}) \\ &\times \prod LR^-(\text{if test} = \text{Neg}) \\ &\times \text{Pre_test Odds} \end{aligned} \quad (3)$$

$$\text{Odds} = \frac{\text{probability}}{1 - \text{probability}} \quad (4)$$

$$LR^+ = \text{Positive Likelihood Ratio} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (5)$$

$$LR^- = \text{Negative Likelihood Ratio} = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \quad (6)$$

$$\text{Certainty Value} = \frac{\text{post test Odds}}{1 + \text{post test Odds}} \quad (7)$$

Using Tree-form Diagnoses

We defined the disease structure in tree form. As described above, the statistical characteristics of a child disease is assumed to be similar to that of the parent disease. If the odds of one diagnosis are updated using the formula presented above, then the odds of the child diagnosis will also be updated. For example, stroke is a child diagnosis of central vertigo, and central vertigo is one child diagnosis of vestibular syndrome. Therefore, if the answer is “No” to the question “Do you see darkness or flashing in front of you?”, then the examination probability of “stroke,” “central vertigo,” and “Vestibular syndrome” is increased. When a specific question such as “Do you have headache or neck pain?” is answered, only the probability of “stroke” might be updated. In this way, one can rank diagnoses with maximum efficiency during an interview.

Preparing to Train from Experts

Forward Calculation

After setting up all these calculation methods, we created 100 simulated patients, and assumed one specific diagnosis for each patient. To simplify a problem, we used six diagnoses that are commonly encountered in emergency departments: presyncope, stroke, brain tumor, vestibular neuritis, Meniere's disease, and benign paroxysmal vertigo. For each patient, we simulated patient responses to interviews using a predetermined sensitivity. Examples of simulated responses to interviews are presented in Table 3. Examples of calculated certainty for the simulated patient are presented in Table 3 and Table 4.

Table 3 – Simulated Answers of a Patient Affected by Benign Paroxysmal Positional Vertigo (BPPV)

Interviews and Answers (Excerpt)

Do you see darkness or flashing in front of you?: NO
Do you have headache or neck pain?: NO
Will the dizziness go well if you keep still?: YES
Did the dizziness occur suddenly?: YES
Does vertigo degenerate with head position conversion?: YES
Are there any initial neurologic symptoms?: NO
Do you drink alcohol every day?: YES
Is your blood pressure high (180 or higher, or 40 higher than usual)?: NO

Table 4 – Calculated Certainty, Differential Diagnoses (top 5)

Certainty value	Differential Diagnosis
0.797	Benign paroxysmal positional vertigo
0.239	Wernicke encephalopathy
0.229	Vestibular neuritis
0.206	Brain tumor
0.182	Central vertigo

Obtaining an Expert Opinion

After defining simulated patients, we produced training data by diagnosing these patients. The computer calculates the post-test probability as a numerical value of 0–1. We can estimate the difference of the probability if we have an accurate value of the post-test probability estimated by physicians. However, physicians do not use numerical probability values in this manner. Therefore, we created a translation table to fill this gap (Table 5).

Table 5 – Certainty Value – Word Translation Table

Word	Certainty Value
Very likely	1
Doubtful	0.8
Possible	0.6
Negative	0.2
Never	0

Using this table, instead of determining the certainty value with numbers, physicians choose words that reflect their degree of confidence from a predetermined word list. If the physician is unable to determine a specific diagnosis, up to two diagnoses to which each certainty word is assigned can be entered. Three physicians evaluated patients and produced diagnoses with certainty using this method. Raters 1 and 2 are emergency physicians. Rater 3 is an otolaryngologist. An input example is presented in Table 6.

Table 6 – Examples of Expert Opinions

Diagnosis	Certainty
Benign paroxysmal positional vertigo	Very likely
Meniere's Disease	Possible

Training from Expert Opinions

For diagnoses for which the certainty value was determined by experts, we calculated the difference of certainty values between a human model and a computer model. In the example presented above, experts claim that the patient is likely to have benign paroxysmal vertigo (BPPV), but believe it might be Meniere's disease. The certainty words are translated into a certainty value using Table 5. We assumed that the difference between a human certainty value and computer certainty value derives from the difference between their assumptions of sensitivity and specificity. The larger the gaps, the greater the correction to the sensitivity and specificity of DtoQ list needs to be. To adapt the computer model to the expert diagnostic model, we used the following steps.

Listing up DtoQ relations to update values

It is impossible for physicians to ascertain certainty values for all differential diagnoses. Therefore, we limited the number of diagnoses a physician can input to two, as described above. The number of diagnoses obtained was so varied that we decided to use only the top diagnoses of their input.

Updating values

We used the extent of the difference between the calculated diagnosis and the human diagnosis as the degree of value update. Let δ be the degree of difference. In the example presented above, the certainty value gap for BPPV between a human and computer is $\delta = 1 - 0.797 = 0.203$, which indicates a slight difference between human knowledge and computer knowledge. In one interview Q1, assuming that this patient answered "assumed answer," then it is necessary to update the 2×2 contingency table of answers (Table 1). We now have gained the knowledge that at least one more patient has a disease and responded with the assumed answer, so we made it a rule to add weight to "a" of Table 1 and then subtracted weight from "b". Letting k be a constant, the contingency table after calculation can be shown as Table 7.

Table 7 – 2×2 Contingency Table After Updating

BPPV to Q1 table		Disease	
		Present	Absent
Interviews	Assumed	$a + k\delta$	$b - k\delta$
	Not assumed	c	d

The DtoQ model we prepared has only sensitivity and specificity values. We must prepare the ratio of patients with disease. To calculate a–d, the ratio of the patients who have the disease must be ascertained using formula (8).

$$\text{Prevalence} = \frac{a+c}{(a+b+c+d)} \quad (8)$$

For this study, we set no specific value of disease prevalence for each disease. We assumed 0.3 for all diseases. It is meaningless to ascertain the absolute value of patients because k can be changed. With this assumption, we were able to calculate a and b respectively using sensitivity and specificity.

After calculating all these values, sensitivity and specificity can be updated using the following formulas (9 and 10).

$$\text{Sensitivity} = \frac{(a+k\delta)}{(a+k\delta)+c} \quad (9)$$

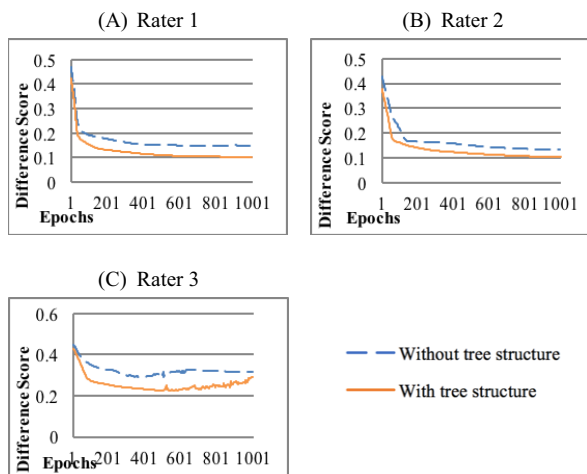
$$\text{Specificity} = \frac{d}{(b-k\delta)+d} \quad (10)$$

Evaluating Models

Differences in the certainty values described above were calculated for physicians' top diagnosis of one patient, and were averaged for all test patients. We designate each as a difference score. The lower this score is, the better the model fits the expert opinions. First, 100 patients were assigned randomly to 80 training sets and 20 test sets, and were averaged for five-fold cross validation. We calculated the scores with each training cycle, which we call epochs. One epoch is equivalent to 80-patient training. Experiment 1: To evaluate the effects of structurizing diagnosis in a tree model, we compared the learning curve by all raters. Experiment 2: To ascertain how this method will function with multiple raters, we evaluated the learning curve by training with multiple raters. In one epoch, the model was trained with 2–3 training data together for each patient. For score calculation, we averaged the difference scores by raters whose data were used to train the model after each epoch. We calculated the inter-rater reliability.

Results

In all experiments, the average difference of certainty values decreased with the number of training cycles. Learning curves of experiment 1 are presented in Fig. 2. By using the tree form in calculation, the score decreased faster. The score became lower with the tree structure for all raters' training data. However, the score fluctuated in later epochs for rater 3.



Figures 2 – Learning Curves of Raters 1, 2 and 3.

In experiment 2, a learning curve with all combinations of raters fluctuated in later epochs (Fig. 3). With a tree structure, a learning curve with raters 1 and 2 showed no tendency to fluctuate until epoch 1000 (Fig. 4).

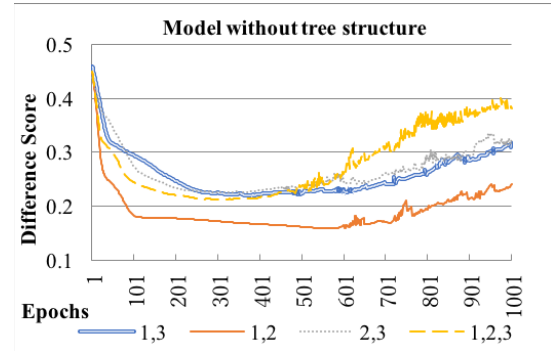


Figure 3 – Learning Curves by Multiple Raters (model without tree structure).

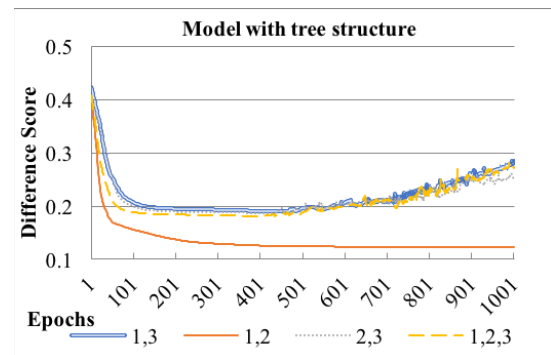


Figure 4 – Learning Curves by Multiple Raters (model with tree structure).

Inter-rater Reliability was poor for raters 1 and 3, and for raters 2 and 3. It was moderate for raters 1 and 2 (Table 8).

Table 8 – Kappa value between raters

κ	Rater 1	Rater 2
Rater 2	0.44	
Rater 3	0.18	0.1

Discussion

Experiment 1 demonstrated that the tree structure of the diagnosis model improves learning speed based on expert opinions. The results fitted better than the model without a tree structure. Experiment 2 showed that if we use training data from raters with different specialties, the training curve fluctuates and increases.

Effects of Tree Structure with a Single Rater

The learning curve shows that the model fits the expert opinion better and faster when one incorporates the tree structure into calculations (Fig. 2A). However, the learning curve shapes varied among raters (Figs. 2B, 2C). The latent disease models of experts that mutually differ might affect the learning curve shape. For rater 3, the curve began to fluctuate from about 350 epochs when we do not incorporate a tree structure, and 500 epochs when we incorporate it. The tree structure of diseases allows updating of many DtoQ relations. Consequently, fitting to the model was done more quickly. Around the epochs of fluctuation, the model appears to be confused by inconsistent training data. The tree structure

might have positive effects on accommodating experts' knowledge.

Effects with Multiple Raters

Experiment 2 showed that training by data from multiple physicians also produces the fluctuation problem (Figs. 3, 4). It did not converge to a specific value for any combination of raters in the model without tree structure (Fig. 3). It also fluctuated with combinations of raters 1 and 3, and raters 2 and 3, and all three raters in the tree model (Fig. 4). However, the score did not fluctuate with a combination of raters 1 and 2 with a tree model within 1000 epochs (Fig. 4). Results show that introduction of the disease tree model seems to stabilize the fluctuation tendency also in training with multiple raters. The model might accept a more inconsistent training set when incorporating a tree structure. Table 8 shows that two physicians having the same specialty (raters 1 and 2, both emergency physicians) have higher inter-rater reliability than physicians between different specialties (raters 1–3, 2–3). The experts' models might be similar if their specialties were the same. The problem of fluctuation seemed to be a result of fitting to inconsistent training data. To understand the characteristics of the fluctuation, we conducted an additional experiment to calculate difference scores from training data instead of test data. Results show that the increase of the score in later epochs also occurs when calculating the difference score with the training data (not presented in this paper). If the model is overfitted, then the difference score will decrease in the training set and will increase in the test set. We were unable to assert that the overfitted model increased the score in later epochs. It remains unclear why the score will increase, but lowering the learning rate with the epoch might relax this problem.

Characteristics of This Method

This method shows some superiority over methods examined in earlier studies. First, the number of parameters, sensitivity, and specificity to each question were less than those of simple Bayesian networks. Fewer training sets are necessary to train the model. Moreover, it becomes easier to manage sensitivity and specificity.

Secondly, the system can learn the DtoQ relation to a partially completed interview. Using this system, one can calculate the diagnosis using a partial answer to the interview. Therefore, we can estimate the error and then update the knowledge base.

Thirdly, the predefined sensitivity and specificity values do not need to be exact. By accumulating training data, one can learn the values of sensitivity and specificity of experts. Eventually, the system will be able to make the same judgments as experts.

Fourth, we can accommodate rare diseases with pre-definition of sensitivity and specificity. Even if the specificity and sensitivity cannot be calculated from the data, one can input clinical knowledge into this system.

Limitations

This study was affected by some limitations. First, the pretest probability was assumed to be equal for all diagnoses. Each disease's probability might differ in different clinical settings and for patients of different age or gender. Therefore, the prevalence of disease in the situation must be input to calculate the post-test probability accurately.

Secondly, the certainty value – word translation table is arbitrary. Some physicians might infer that a diagnosis is “very likely” for a certain patient, whereas other physicians might regard it as “doubtful,” even if they regard the diagnosis to be valid to the same degree.

Thirdly, the number of features to update for a given training set might be insufficient. The physician might have made decisions using information that was not included in the model. The low number of raters is also noteworthy. In future works, we plan to clarify the relations between fluctuation phenomena, model overfitting, and inconsistent training data. Many points of system improvement remain. Nevertheless, this prototype is expected to play an important role in elucidating system problems.

Conclusion

For this study, we assessed effects of integrating a disease tree structure into a Bayesian model for calculating post-test probability. The tree structure improved the learning rate without increasing the number of model parameters. Results also show that inconsistent training data tend to affect the model adversely. Future works will investigate means of stratifying and estimating prevalence and prior probabilities.

This paper presented an approach to acquire specialists' latent knowledge and to develop clinical reasoning model. This approach can be useful for developing consultation systems that will be helpful for situations in which access to specialists is restricted.

Acknowledgements

The authors declare that they have no conflict of interest associated with this study or this report describing it. We gratefully acknowledge the rating efforts of two raters.

References

- [1] V.L. Patel, E.H. Shortliffe, M. Stefanelli, P. Szolovits, M.R. Berthold, R. Bellazzi, and A. Abu-Hanna, The coming of age of artificial intelligence in medicine, *Artif Intell Med* **46** (2009), 5-17.
- [2] J.H. Chen, T. Podchiyska, and R.B. Altman, OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records, *J Am Med Inform Assoc* **23** (2016), 339-348.
- [3] R.A. Miller, H.E. Pople, Jr., and J.D. Myers, Internist-1, an experimental computer-based diagnostic consultant for general internal medicine, *N Engl J Med* **307** (1982), 468-476.
- [4] D.F. Sittig, A. Wright, J.A. Osherooff, B. Middleton, J.M. Teich, J.S. Ash, E. Campbell, and D.W. Bates, Grand challenges in clinical decision support, *J Biomed Inform* **41** (2008), 387-392.
- [5] D.E. Heckerman, and R.A. Miller, Towards a better understanding of the INTERNIST-1 knowledge base, *Medinfo* **86** (1986), 27-31.
- [6] S.G. Pauker, and J.P. Kassirer, The threshold approach to clinical decision making, *N Engl J Med* **302** (1980), 1109-1117.
- [7] A. Shahrami, M. Norouzi, H. Kariman, H.R. Hatamabadi, and A. Arhami Dolatabadi, True Vertigo Patients in Emergency Department: an Epidemiologic Study, *Emerg (Tehran)* **4** (2016), 25-28.

Address for correspondence

Satoshi Iwai, M.D.

Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655 Japan

+81-3-5800-6427

stihahi@gmail.com

A Consensus-Based Approach for Harmonizing the OHDSI Common Data Model with HL7 FHIR

Guoqian Jiang^a, Richard C. Kiefer^a, Deepak K. Sharma^a, Eric Prud'hommeaux^b, Harold R. Solbrig^a

^a Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

^b W3C/MIT, Boston, MA, USA

Abstract

A variety of data models have been developed to provide a standardized data interface that supports organizing clinical research data into a standard structure for building the integrated data repositories. HL7 Fast Healthcare Interoperability Resources (FHIR) is emerging as a next generation standards framework for facilitating health care and electronic health records-based data exchange. The objective of the study was to design and assess a consensus-based approach for harmonizing the OHDSI CDM with HL7 FHIR. We leverage a FHIR W5 (Who, What, When, Where, and Why) Classification System for designing the harmonization approaches and assess their utility in achieving the consensus among curators using a standard inter-rater agreement measure. Moderate agreement was achieved for the model-level harmonization ($\kappa=0.50$) whereas only fair agreement was achieved for the property-level harmonization ($\kappa=0.21$). FHIR W5 is a useful tool in designing the harmonization approaches between data models and FHIR, and facilitating the consensus achievement.

Keywords:

Reference standards; Observational study; Vocabulary, controlled

Introduction

Integrated Data Repositories (IDRs) [1-2] are needed to combine molecular and phenotypic data, making data available for use with analytic tools. This is especially important for clinical research investigators with limited computing resources. In a 2010 survey conducted by the Clinical and Translational Science Award (CTSA) consortium [3], IDR was defined as a data warehouse integrating various sources of clinical data to support queries for a range of research-like functions. Survey results suggest that individual organizations are progressing in their approaches to the development, management, and use of IDRs as a means to support a broad array of research. A variety of data models have been developed to provide a standardized data interface that supports organizing clinical research data into a standard structure in such IDRs. These data models include the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [4-5], the National Patient-Centered Research Networks (PCORnet) CDM [6], and the Informatics for Integrating Biology and the Bedside (i2b2) Star Schema [7].

OMOP was a public-private partnership established to inform the appropriate use of observational healthcare databases for

studying the effects of medical products. The OMOP community is actively using the OMOP CDM [4-5] for their various research purposes. Observational Health Data Sciences and Informatics (OHDSI) has been established as a multi-stakeholder, interdisciplinary collaborative to create open-source solutions that bring out the value of observational health data through large-scale analytics. The OHDSI collaborative includes all of the original OMOP research investigators and develops its tools using the OMOP CDM, which will continue to be an open-source, community standard for observational healthcare data.

The PCORnet CDM [6] is based on the Mini-Sentinel CDM and has been informed by other distributed initiatives such as the HMO Research network, the Vaccine Safety Datalink, and the ONC Standards & Interoperability Framework Query Health Initiative. The CDM leverages standard terminologies and coding systems for healthcare to enable interoperability and ensure responsiveness to evolving data standards.

The i2b2 is an open-source clinical data analytics platform that provides a component-based architecture and a flexible analytical database design [7-9]. The i2b2 Star Schema was developed as a CDM that enables conformant transformation of patient data to a common data structure and representation of meaning. i2b2-based solutions have been widely used in clinical research communities such as the Shared Health Research Information Networks (SHRINE) [10], and the PCORnet [11]. Building on the i2b2 framework, the tranSMART platform [12-14] is an analytic platform that also incorporates the ability to load molecular datatypes, including those derived from next generation sequencing (NGS).

These data models serve well as a layer of standardization for clinical research data within their own research network; however, if the investigators want to reuse and integrate these research datasets and applications in broader clinical research communities across different research networks they still face huge challenges. This situation demands a global data model as a reference standard to facilitate data model harmonization and data integration.

HL7 Fast Healthcare Interoperability Resources (FHIR) is emerging as a next generation standards framework for facilitating health care and electronic health records-based data exchange [15]. However, it has been a challenging issue for the standards and research communities to build a consistent and measurable approach for enabling the consensus achieving in terms of data model harmonization efforts. The objective of the study is to design and assess a consensus-based approach for harmonizing the OHDSI CDM with HL7 FHIR. We leverage the FHIR W5 (Who, What,

When, Where, and Why) Classification System [16-17] for designing the harmonization approach and assess its utility in achieving the consensus among curators using a standard inter-rater agreement measure. The outcome of this study would provide guidance to harmonize different data models with FHIR in future.

Methods

Materials

OMOP CDM

The OMOP CDM is “designed to support the conduct of research to identify and evaluate associations between interventions (e.g., drug exposure) and outcomes (e.g., adverse effects) caused by these interventions” [5]. The design principles of the CDM include: 1) suitable for purpose; 2) data protection; 3) design of domains; 4) rational for domains; 5) standardized vocabularies; 6) reuse of existing vocabularies; 7) maintaining source codes; 8) technology neutrality; 9) scalability; and 10) backwards compatibility. The CDM defines table schemas in a person-centric manner. As of September 18, 2016, version V5.0.1 of the CDM was released, which contains 39 tables in 6 categories: standardized clinical data, standardized health system data, standardized health economics, standardized metadata, standardized vocabularies and standardized derived elements. In fact, terminology normalization enabled by standard vocabularies with focus on SNOMED CT, LOINC and RxNorm is a strong characteristic of the OMOP CDM. Figure 1 shows a diagram highlighting the high-level relationships among the tables and categories.

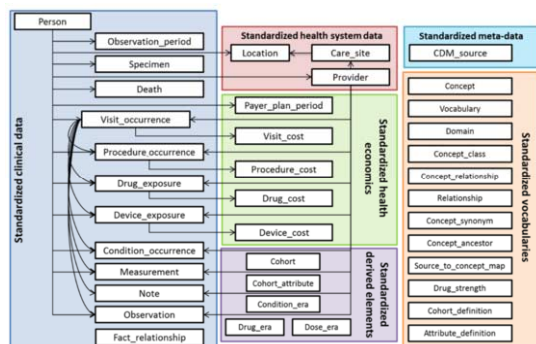


Figure 1 – A diagram highlighting the high-level relationships among the tables and categories. (Source from the OMOP CDM document)

HL7 FHIR Core Resources

Detailed Clinical Models (DCMs) have been regarded as the basis for retaining computable meaning when data are exchanged between heterogeneous computer systems [18-19]. Amongst the emerging national and international initiatives on the standardization of DCM modeling are the Clinical Informatics Modeling Initiative (CIMI) [20] and FHIR [14]. The primary goal of CIMI is to “Improve the interoperability of healthcare systems through shared implementable clinical information models (A single curated collection)”. FHIR builds around the concept of “resources”. “Resources” here means small discrete concepts with clearly defined scope that can be maintained independently. Resources are the smallest units of a transaction, and each resource has a unique id that aligns with RESTful design philosophy. As of September 16,

2016, the version Draft Standard for Trials Use (DSTU) 2 has been released and the version of STU 3 was placed under ballot. Figure 2 shows a collection of FHIR core resources under the FHIR Clinical category in the released DSTU 2.

HL7 FHIR W5 Classification System

The HL7 electronic health record (EHR) workgroup develops a FHIR Record Lifecycle Events Implementation Guide that “offers a methodology to support trusted EHR management using HL7 FHIR” [17]. The fundamental assumption is that the EHR system captures Action taken and creates corresponding Record Entries. Actions have associated metadata (e.g., who, what, when, where, why, how, under what conditions, in what context). The corresponding Record Entry captures this metadata along with other Action and Record Entry related information. The FHIR W5 (Who, What, When, Where, and Why) Report [16] has been created for each release to establish consistent W5 metadata definitions across FHIR resources for tracking resource (instance) lifecycle events when content is created or updated, and/or when signature is applied. In the current implementation, the FHIR W5 Classification System plays two roles: 1) specifying a Resource Category system to classify the FHIR resources; and 2) specifying a Property Category system to classify the FHIR properties. For example, as shown in Figure 2, the FHIR resource Observation belongs to the category clinical.diagnostics. Table 1 and Table 2 show the categories specified in both systems respectively and examples from FHIR W5 Report for the properties of the FHIR resource Observation.

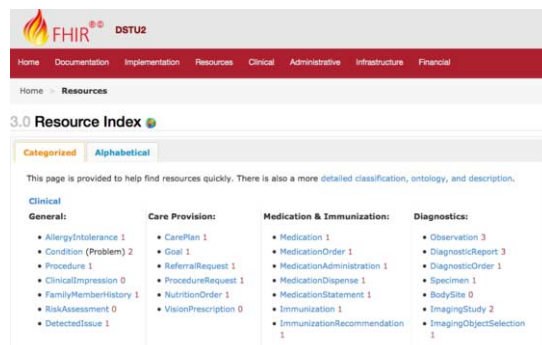


Figure 2 – A screenshot showing a collection of FHIR core resources under the Clinical category in the released DSTU 2.

Measure the Consensus in the Model-Level Harmonization

The OMOP CDM consists of a collection of table schemas. Each table schema represents a particular OMOP domain, which is analogous to the resources as defined in FHIR. In the model level, we design the following two approaches to align the table schemas with the FHIR resources. The first approach is to categorize the OMOP table schemas using the FHIR W5 resource categories, and the second approach is to map the OMOP table schemas directly with the FHIR core resources.

For the first approach, we extracted all table names and their textual descriptions from the OMOP CDM documentation website [5]. We designed a mapping application using an Excel spreadsheet that allows curators to assign a FHIR W5 resource category to each OMOP table name. As a pilot study, we asked the project team members to use the mapping application to assign the FHIR W5 resource category independently. The purposes of the pilot study are two-fold: 1) to examine the inter-rater agreement for categorizing the

For the second approach, we extracted the OMOP table names and their descriptions under the category of “standardized clinical data” [5] and the FHIR core resource names and their definitions from the DSTU2 version. We also designed the Excel spreadsheet-based mapping application. We used four SKOS mapping properties (exactMatch, closeMatch, broadMatch, and narrowMatch) as the mapping types. As a pilot study, we asked each project team member to use the mapping application to map the OMOP table schemas under the category of “standardized clinical data” to the FHIR core resources individually. The purpose of the pilot study is to examine the inter-rater agreement for creating the mappings.

	Topic area	Description	Field or Category	Notes
		The PERSON data contains records that uniquely identify each patient in the source data who has time at risk for clinical events recorded within the		
PERSON				
		The OBSERVATION_PERIOD data contains records which uniquely define the time of when for which a Person is at risk for clinical events recorded within the source data, except the events at risk for recording healthy patients with no healthcare encounters.		
OBSERVATION_PERIOD				administrative administrative.device administrative.location administrative.provider administrative.provider.organization
		The SPECIMEN data contains records that are identified through sample number and location.		
SPECIMEN				
		The DEATH data contains the clinical event at which time a Person dies. A person can have as many as one record if the source data contains records without the event.		
DEATH				clinical clinical.general clinical.general.organization clinical.general.provider
		The VISIT_OCCURRENCE data contains the dates of time a person continually receives medical services from an or more providers in relation to a given patient within the health care system. Visits can include office visits, inpatient care, treatment coordination, emergency, and long term care. Visits may transition between these settings over the course of the patient's care.		conformances conformances.behavior conformances.conformances conformances.conformances.conformances
VISIT_OCCURRENCE				conformances.terminology conformances.terminology.conformances
		The PROCEDURE_OCCURRENCE data contains records of all procedures performed by a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient. The procedure data contains records of all procedures performed on a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient. The procedure data contains records of all procedures performed on a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient.		financial financial.knowledge financial.knowledge.conformances financial.knowledge.conformances.conformances
PROCEDURE_OCCURRENCE				financial.knowledge.conformances.terminology financial.knowledge.conformances.terminology.conformances
		The DRUG_EXPOSURE data contains records of all drugs administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient. The drug exposure data contains records of all drugs administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient.		infrastructure.documents infrastructure.documents.conformances infrastructure.documents.conformances.conformances
DRUG_EXPOSURE				infrastructure.documents.conformances.terminology infrastructure.documents.conformances.terminology.conformances
		The DEVICE_EXPOSURE data contains records of all devices administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient. The device exposure data contains records of all devices administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient.		infrastructure.documents.conformances.terminology infrastructure.documents.conformances.terminology.conformances
DEVICE_EXPOSURE				infrastructure.documents.conformances.terminology.conformances
		The CONDITION_OCCURRENCE data contains records of all conditions administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient. The condition occurrence data contains records of all conditions administered to a patient, carried out, a healthcare provider on the patient by a date, a diagnosis, a procedure, a device, a provider, a location, and a patient.		infrastructure.documents.conformances.terminology.conformances
CONDITION_OCCURRENCE				infrastructure.documents.conformances.terminology.conformances

Measure the Consensus in the Property-Level Harmonization

A	B	C	D	E	F	G
Field	Required	Type	Description	enum or category	Notes	
observation_id		Integer	A unique identifier for each observation.			
species_id	Yes	Integer	A foreign key identifier to the Species table where the observation was recorded. The foreign key value of this field and the value of the observation_id field are used to identify the observation.	class species		
observer	Yes	Text	The name of the observer who recorded the observation.	enum observer		
observation_type_name	Yes	Text	A foreign key to the observation_type_name table in the Standardized Observations table. The value of this field and the value of the observation_id field are used to identify the observation.	enum observation_type_name		
value_number	Yes	Float	The observation record assigned a number. This is applicable to observations where the value is a number or a text value.			
value_string	Yes	Text	The observation record assigned a string. This is applicable to observations where the value is a string or a text value.			
value_date	Yes	Date	The observation record assigned a date. This is applicable to observations where the value is a date or a text value.			
value_time	Yes	Time	The observation record assigned a time. This is applicable to observations where the value is a time or a text value.			
value_boolean	Yes	Boolean	The observation record assigned a boolean value. This is applicable to observations where the value is a boolean or a text value.			
value_integer	Yes	Integer	The observation record assigned an integer value. This is applicable to observations where the value is an integer or a text value.			
value_float	Yes	Float	The observation record assigned a float value. This is applicable to observations where the value is a float or a text value.			
value_text	Yes	Text	The observation record assigned a text value. This is applicable to observations where the value is a text or a text value.			
value_date_time	Yes	Date Time	The observation record assigned a date and time value. This is applicable to observations where the value is a date and time or a text value.			
value_boolean_text	Yes	Boolean Text	The observation record assigned a boolean and text value. This is applicable to observations where the value is a boolean and text or a text value.			
value_integer_text	Yes	Integer Text	The observation record assigned an integer and text value. This is applicable to observations where the value is an integer and text or a text value.			
value_float_text	Yes	Float Text	The observation record assigned a float and text value. This is applicable to observations where the value is a float and text or a text value.			
value_date_text	Yes	Date Text	The observation record assigned a date and text value. This is applicable to observations where the value is a date and text or a text value.			
value_time_text	Yes	Time Text	The observation record assigned a time and text value. This is applicable to observations where the value is a time and text or a text value.			
value_boolean_integer	Yes	Boolean Integer	The observation record assigned a boolean and integer value. This is applicable to observations where the value is a boolean and integer or a text value.			
value_boolean_float	Yes	Boolean Float	The observation record assigned a boolean and float value. This is applicable to observations where the value is a boolean and float or a text value.			
value_boolean_date	Yes	Boolean Date	The observation record assigned a boolean and date value. This is applicable to observations where the value is a boolean and date or a text value.			
value_boolean_time	Yes	Boolean Time	The observation record assigned a boolean and time value. This is applicable to observations where the value is a boolean and time or a text value.			
value_boolean_date_time	Yes	Boolean Date Time	The observation record assigned a boolean and date and time value. This is applicable to observations where the value is a boolean and date and time or a text value.			

Inter-Rater Agreement Analysis

over that which would be expected by chance. The Kappa statistics value of 1 ($k = 1$) reflects complete agreement among raters and, a value of zero or less ($k \leq 0$) shows no agreement.

Results

Five project team members were asked to complete the mapping spreadsheets and four valid responses were received. The average time used for the FHIR W5 Resource Category approach was 26 minutes (range from 24-30 minutes) whereas the average time used for the direct mapping approach was 29.25 minutes (range 17-35 minutes).

The overall Fleiss' kappa statistics for categorizing the OMOP CDM tables with the FHIR W5 Resource Classification System was calculated as 0.50 whereas the overall Fleiss' kappa statistics for directly mapping the OMOP CDM tables with the FHIR core resources was calculated as 0.48. The results indicated that for both harmonization approaches, the team achieved moderate agreement.

After we calculated the Fleiss's kappa statistics, we had a team discussion to resolve the disagreement and achieved the consensus. Table 1 shows the distribution of the FHIR core resources and OMOP CDM tables by the FHIR W5 resource category.

FHIR W5 Resource Category	Number of FHIR Resources (DSU2)	Number of OMOP CDM Tables
administrative	subtotal: 16 (17%)	subtotal: 7 (18%)
administrative.device	3	1
administrative.entity	4	1
administrative.group	3	4
administrative.individual	6	1
clinical	subtotal: 27 (28%)	subtotal: 13 (33%)
clinical.careprovision	6	
clinical.diagnostics	7	6
clinical.general	7	3
clinical.medication	7	4
conformance	subtotal: 10 (10%)	subtotal: 9 (23%)
conformance.behavior	3	
conformance.content	2	
conformance.misc	2	
conformance.terminology	3	9
financial	subtotal: 10 (10%)	subtotal: 5 (13%)
financial.billing	2	4
financial.other	1	
financial.payment	2	1
financial.support	5	
infrastructure	subtotal: 16 (17%)	subtotal: 3 (8%)
infrastructure.documents	4	2
infrastructure.exchange	4	1
infrastructure.information	4	
infrastructure.structure	4	
workflow	subtotal: 17 (18%)	subtotal: 2 (5%)
workflow.encounter	4	2
workflow.order	9	
workflow.scheduling	4	
Total	96 (100%)	39 (100%)

The results indicated that the OMOP CDM covers the domains in all 6 categories as defined by FHIR W5. The distribution of the OMOP CDM tables in the category of clinical (33%), financial (13%) and administrative (18%) is comparable with FHIR. The percentage of the OMOP CDM

tables in the category of infrastructure (8%) and workflow (5%) is lower than that in FHIR whereas the percentage of the OMOP CDM tables in the category of conformance (23%) is higher than that in FHIR. Interestingly, the OMOP CDM tables in the category conformance are all classified in the sub-category *conformance.terminology*, highlighting the OMOP CDM design principles emphasizing the use of standardized vocabularies.

Property-Level Harmonization: Inter-Rater Agreement Analysis for the FHIR W5 Property Classification System

Similarly, five project team members were asked to complete the mapping spreadsheets and four valid responses were received. The average time used for the OMOP *Observation* table was 15.75 minutes (range 8-30 minutes) and the average time used for the OMOP *Drug_Exposure* was 16.75 minutes (range from 7-30 minutes).

The overall Fleiss' kappa statistics for categorizing the fields in the OMOP CDM tables "*Observation*" and "*Drug_Exposure*" with the FHIR W5 Property Classification System were both calculated as 0.21. The results indicated that for the property categorization approach, the team only achieved fair agreement. One of the main reasons is because that the FHIR W5 Property Classification System is new to most of the team members and only a limited documentation is available. We anticipate that more training orientation to curators would help improve the inter-rater agreement.

Property-Level Harmonization: Alignment between the FHIR Resource Elements and the OMOP Table Fields

We aligned the fields of the two OMOP tables *Observation* and *Drug_Exposure* with the elements from the FHIR resource *Observation* and *DrugAdministration* (based on the FHIR W5 Report [16]¹ based on the FHIR W5 Property Classification System. The alignment results as shown in Table 2 indicated that a number of property categories including *class*, *context*, *identifier*, *when.done*, *where*, *who.actor*, *who.focus* are aligned reasonably well. A list of fields in both OMOP tables were categorized in the property category "*what*", indicating this "*what*" category may need to be refined to have more specific subcategories.

Table 2 – The alignment results between the FHIR resource elements and the OMOP table fields.

FHIR W5 Property Category	FHIR Observation	OMOP Observation	FHIR DrugAdministration	OMOP Drug_Exposure
class	category	observation_type_concept_id		drug_type_concept_id
context	encounter	visit_occurrence_id	encounter	visit_occurrence_id
grade		qualifier_concept_id		
id	identifier	observation_id	identifier	drug_exposure_id
id.version				
status	status		status	
what	code	observation_concept_id; value_as_number value_as_string value_as_concept_id unit_concept_id observation_source_value observation_source_concept_id unit_source_value qualifier_source_value		drug_concept_id refills quantity days_supply sig route_concept_id effective_dose dose_unit_concept_id lot_number drug_source_value drug_source_concept_id route_source_value dose_unit_source_value
when.done	effectiveTime[x]	observation_date	effectiveTime[x]	drug_exposure_end_date
when.init		observation_time		drug_exposure_start_date
when.planned				
when.recorded	issued			
where		visit_occurrence_id		visit_occurrence_id
who				
who.actor	performer	provider_id	practitioner	provider_id
who.author				
who.cause				
who.focus	subject	person_id	patient	person_id
who.source				
who.witness				
why				stop_reason

¹ Note that not all properties of the FHIR resources are categorized in the FHIR W5 Report.

Discussion

The use of data integration standards plays a critical role in the increasing adoption of the OHDSI CDM-based data repositories for clinical observational studies [22]. While the OHDSI Vocabulary CDM [23] and its vocabulary services [24] have provided a solid foundation for enabling semantic interoperability across different clinical and research systems, the heterogeneous data model use in these systems remains a major barrier for data integration and sharing. The emerging HL7 FHIR aims to serve as a global reference standard for exchanging healthcare and EHR data, and mappings to FHIR from different data models would greatly facilitate the secondary use of EHR data for clinical and translation research including the observational studies. Therefore, a number of independent efforts in harmonizing between FHIR and OHDSI CDM are currently underway. For example, Choi and Duke's team has developed a preliminary prototype known as the OHDSI on FHIR platform with OHDSI CDM mappings to FHIR resources [25]. Actually, the FHIR Infrastructure Work Group is developing a Data Access Framework (DAF) FHIR Implementation Guide which includes the guidance on creating profiles and data element mappings between FHIR and OMOP CDM (mainly in the model level) [26]. However, consensus building is a critical yet challenging factor in harmonizing and standardizing the mappings between different data models. Our consensus-based approach using the FHIR W5 category system is a preliminary effort in helping achieve the community-based agreement in a consistent and measurable manner.

Conclusions

In this study, we designed and assessed a consensus-based approach for harmonizing the OMOP CDM with HL7 FHIR. We leveraged a FHIR W5 (Who, What, When, Where, and Why) Classification System for designing the harmonization approaches and assessed their utility in achieving the consensus among curators using a standard inter-rater agreement measure. We demonstrated that the FHIR W5 classification system is a useful and promising tool for designing the harmonization approach between a data model and FHIR, and for facilitating the consensus achievement among curators. Future work includes: 1) building and refining the FHIR W5 Ontology using a community-based approach; 2) mapping the FHIR W5 ontology with other real-world data models (e.g., PCORnet CDM, i2b2); and 3) enhancing the mapping applications for effectively facilitating the alignment between various data models and FHIR.

Acknowledgements

This study is supported in part by NIH grants U01 HG009450, U01 CA180940, and R01 GM105688.

References

- [1] V. Huser and J.J. Cimino, Desiderata for healthcare integrated data repositories based on architectural comparison of three public repositories, *AMIA Annu Symp Proc* **2013** (2013), 648-656.
- [2] T.D. Wade, P.T. Zelarney, R.C. Hum, S. McGee, and D.H. Batson, Using patient lists to add value to integrated data repositories, *J Biomed Inform* **52** (2014), 72-77.
- [3] S.L. MacKenzie, M.C. Wyatt, R. Schuff, J.D. Tenenbaum, and N. Anderson, Practices and perspectives on building integrated data repositories: results from a 2010 CTSA survey, *J Am Med Inform Assoc* **19** (2012), e119-124.

- [4] The OMOP Common Data Model, in, 2015.
- [5] OMOP Common Data Model V5.0.1, in, 2016.
- [6] S.R. Partridge, K. Balestracci, A.T. Wong, L. Hebden, K. McGeechan, E. Denney-Wilson, M.F. Harris, P. Phongsavan, A. Bauman, and M. Allman-Farinelli, Effective Strategies to Recruit Young Adults Into the TXT2BfIT mHealth Randomized Controlled Trial for Weight Gain Prevention, *JMIR Res Protoc* **4** (2015), e66.
- [7] D.M. Lopez and B. Blobel, mHealth in Low- and Middle-Income Countries: Status, Requirements and Strategies, *Stud Health Technol Inform* **211** (2015), 79-87.
- [8] J.G. Klann and S.N. Murphy, Computing health quality measures using Informatics for Integrating Biology and the Bedside, *J Med Internet Res* **15** (2013), e75.
- [9] J.G. Klann, M.D. Buck, J. Brown, M. Hadley, R. Elmore, G.M. Weber, and S.N. Murphy, Query Health: standards-based, cross-platform population health surveillance, *J Am Med Inform Assoc* **21** (2014), 650-656.
- [10] G.M. Weber, S.N. Murphy, A.J. McMurry, D. Macfadden, D.J. Nigrin, S. Churchill, and I.S. Kohane, The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories, *J Am Med Inform Assoc* **16** (2009), 624-630.
- [11] S.E. Daugherty, S. Wahba, and R. Fleurence, Patient-powered research networks: building capacity for conducting patient-centered clinical outcomes research, *J Am Med Inform Assoc* **21** (2014), 583-586.
- [12] B.D. Athey, M. Braxenthaler, M. Haas, and Y. Guo, transSMART: An Open Source and Community-Driven Informatics and Data Sharing Platform for Clinical and Translational Research, *AMIA Jt Summits Transl Sci Proc* **2013** (2013), 6-8.
- [13] E. Scheufele, D. Aronson, R. Coopersmith, M.T. McDuffie, M. Kapoor, C.A. Uhrich, J.E. Avitabile, J. Liu, D. Housman, and M.B. Palchuk, transSMART: An Open Source Knowledge Management and High Content Data Analytics Platform, *AMIA Jt Summits Transl Sci Proc* **2014** (2014), 96-101.
- [14] A.J. Novak, Y.W. Asmann, M.J. Maurer, C. Wang, S.L. Slager, L.S. Hodge, M. Manske, T. Price-Troska, Z.Z. Yang, M.T. Zimmermann, G.S. Nowakowski, S.M. Ansell, T.E. Witzig, E. McPhail, R. Ketterling, A.L. Feldman, A. Dogan, B.K. Link, T.M. Habermann, and J.R. Cerhan, Whole-exome analysis reveals novel somatic genomic alterations associated with outcome in immunochemotherapy-treated diffuse large B-cell lymphoma, *Blood cancer journal* **5** (2015), e346.
- [15] M.S. Clemens, T.A. Heafner, J.D. Watson, J.K. Aden, 3rd, T.E. Rasmussen, and S.C. Glasgow, Quality of Life in United States Veterans With Combat-Related Ostomies From Iraq and Afghanistan, *Mil Med* **181** (2016), e1569-e1574.
- [16] FHIR W5 Report, in, 2016.
- [17] L.R. Greenwood, M.G. Cantaloub, J.L. Burnett, A.W. Myers, C.T. Overman, J.B. Forrester, B.G. Glasgow, and H.S. Miley, Low-background gamma-ray spectrometry for the international monitoring system, *Appl Radiat Isot* (2016).
- [18] J.F. Coyle, A.R. Mori, and S.M. Huff, Standards for detailed clinical models as the basis for medical data exchange and decision support, *Int J Med Inform* **69** (2003), 157-174.
- [19] G. Jiang, J. Evans, T.A. Oniki, J.F. Coyle, L. Bain, S.M. Huff, R.D. Kush, and C.G. Chute, Harmonization of detailed clinical models with clinical study data standards, *Methods Inf Med* **54** (2015), 65-74.
- [20] E.R. Gamazon and B.E. Stranger, The impact of human copy number variation on gene expression, *Briefings in functional genomics* **14** (2015), 352-357.
- [21] Fleiss' kappa, in, 2016.
- [22] G. Hripsak, P.B. Ryan, J.D. Duke, N.H. Shah, R.W. Park, V. Huser, M.A. Suchard, M.J. Schuemie, F.J. DeFalco, A. Perotte, J.M. Banda, C.G. Reich, L.M. Schilling, M.E. Matheny, D. Meeker, N. Pratt, and D. Madigan, Characterizing treatment pathways at scale using the OHDSI network, *Proc Natl Acad Sci U S A* **113** (2016), 7329-7336.
- [23] OHDSI Vocabulary CDM, in, 2016.
- [24] E.D. Spikol, C.E. Laverriere, M. Robnett, G. Carter, E. Wolfe, and E. Glasgow, Zebrafish Models of Prader-Willi Syndrome: Fast Track to Pharmacotherapeutics, *Diseases* **4** (2016).
- [25] M. Choi, R. Starr, M. Braunstein, and J.D. Duke, OHDSI on FHIR platform development with OMOP CDM mapping to FHIR resources, in, 2016.
- [26] DAF-Research Profile List and Mappings from FHIR to PCORnet CDM and OMOP CDM, in, 2017.

Address for correspondence

Guoqian Jiang, MD, PhD at jiang.guoqian@mayo.edu.

Text Data Mining of Aged Care Accreditation Reports to Identify Risk Factors in Medication Management in Australian Residential Aged Care Homes

Tao Jiang^a, Siyu Qian^a, David Hailey^a, Jun Ma^b, Ping Yu^a

^a Centre for IT-enabled Transformation, School of Computing and Information Technology,
University of Wollongong, Wollongong, NSW, Australia

^b SMART Infrastructure Facility, University of Wollongong, Wollongong, NSW, Australia

Abstract

This study aimed to identify risk factors in medication management in Australian residential aged care (RAC) homes. Only 18 out of 3,607 RAC homes failed aged care accreditation standard in medication management between 7th March 2011 and 25th March 2015. Text data mining methods were used to analyse the reasons for failure. This led to the identification of 21 risk indicators for an RAC home to fail in medication management. These indicators were further grouped into ten themes. They are overall medication management, medication assessment, ordering, dispensing, storage, stock and disposal, administration, incident report, monitoring, staff and resident satisfaction. The top three risk factors are: “ineffective monitoring process” (18 homes), “noncompliance with professional standards and guidelines” (15 homes), and “resident dissatisfaction with overall medication management” (10 homes).

Keywords:

Risk Factors; Medication Errors; Nursing Homes

Introduction

With population aging, the demand for aged care services around the world is increasing. Associated with the aging process is an increased level of frailty and chronic diseases, which pose major challenges to RAC services [1]. Residential aged care homes in Australia provide accommodation, nursing care and personal care services for the frail older people [2]. Previous research suggests that people living in RAC homes have a higher exposure to various risk factors than their counterparts in the community [1]. Long-term and complex chronic conditions associated with the aging process are the main challenges for nursing staff to provide appropriate care to these people [2]. The high demands for appropriate care and regulatory compliance have led to high cost and burden for aged care services [3]. Formal aged care services in Australia are predominantly financed by taxpayers with some contributions from service users [4]. In order to protect residents' safety and enhance the quality of the services, the Australian government has imposed stringent accreditation and safety standards through its aged care accreditation program administered by Australian government Aged Care Quality Agency (ACQA).

The aged care accreditation program in Australia focuses on continuous quality improvement strategies [5]. It includes an accreditation process and monitoring of ongoing performance against standards [5]. It is an effective approach to risk management and quality improvement of government-subsidized RAC homes [4]. In Australia, RAC homes are

required to meet the accreditation standards at all times to ensure a high standard of care and services [6]. If a home fails in the accreditation, a timetable for improvement with a deadline will be developed by the ACQA [7]. Meanwhile, the agency monitors the home's progress in making improvements. If the home does not meet all the requirements before the deadline, the agency may conduct a review audit which may result in the home's accreditation certification being revoked and the home will lose legibility for receiving government subsidy.

The aged care accreditation system is established to manage potential risks in RAC homes. The accreditation process starts with self-assessment, followed by a desk audit, a site audit, the decision whether or not to accredit the home and the publication of the accreditation report [4].

Self-assessment, desk audit and site audit can help the RAC homes to identify the risk areas and risk factors [7]. The accreditation teams' findings, the decision of the accreditation agency about whether an RAC home has met the 44 expected outcomes is the official verdict about the home's risk management system. Therefore, a risk management approach is essential for RAC homes to pass the accreditation [7].

The whole process of risk management includes identifying risks, assessing the risks, developing risk management plans, implementing risk management actions and re-evaluating risks which have occurred in the process of delivering aged care services [7]. A vulnerable area where risks for resident safety might occur in RAC homes is medication management.

Residential aged care homes must support and safely manage each resident's medication need [8]. It is reported that residents take an average of seven to nine medications [9-12]. As medication management is a complex process involving prescription, ordering, delivery, administration, monitoring, evaluation and documentation [13], errors such as wrong drug that are detrimental to medication safety may occur in any stage [14]. The error rate is between 28% and 40% [15, 16]. The occurrence of these errors may be increased by nurses' high physical and mental load [17] and large amount of medication to be administered under time pressures [18]. Therefore, this research aimed to identify the risk factors related to medication management in RAC homes.

Methods

We followed a three-step process to extract the data from the aged care accreditation reports in Australia: (1) report collection, (2) section/paragraph extraction, and (3) key-words/terminology identification (see Figure 1).

Table 1 - Risk factors for RAC homes to fail in accreditation standard of medication management

Issue	Risk factors	No. of homes
Overall medication management	Non compliance	15
	Inconsistent policies	2
	Irregular medication review	1
Assessment not done	Medication are not assessed	4
Medication ordering	Non compliance with medical officer's directives	1
Dispensing	No pre-packed medication system	1
Medication stocking	Insufficient or unavailable stock	2
	Insecure storage and disposal	5
Medication administration	Incomplete information	3
	Inconsistent timing and practice	8
	Use of controlled medication for other residents	1
	Use of unordered medication	2
Monitoring	Ineffective process	18
	Staff practice not monitored	4
Incidents	Incident not identified and acted upon	3
	Inconsistent follow up	2
Staff satisfaction	Lack of competent assessment	3
	Insufficient staffing	1
	Lack of education	1
Resident satisfaction	Dissatisfied with medication management	10
	Dissatisfied with medication administration	2

There were two risk factors in medication stocking. Two RAC homes did not have adequate stock or medications ran out. Five RAC homes did not store and dispose medications safely and securely.

There were four risk factors in the medication administration process: (1) medication charts did not contain complete and sufficient information in three RAC homes; (2) in eight RAC homes, medications were not administered, consistently safely and correctly, e.g. at appropriate intervals, not following prescribed orders, or inconsistent administration process; (3) in one RAC home, controlled medications prescribed for individual residents were used for other residents; and (4) in two RAC homes, medications being administered were not ordered by a medical officer.

The two risk factors for monitoring included (1) monitoring processes were not in place or were not effective; which happened in all failed RAC homes; and (2) four RAC homes did not monitor staff practice.

There were two risk factors for incident reporting: (1) medication incidents were not identified, analysed and acted upon (2 RAC homes); (2) medication incidents were not consistently followed up with the relevant parties (2 RAC homes).

There were three factors for staff satisfaction. These included (1) not conducting competency assessment (3 RAC homes); (2) insufficient staffing (1 RAC home); and (3) not providing education to staff in relation to medications (1 RAC home).

There were two risk factors for resident dissatisfaction. Residents were either not satisfied with medication management (10 RAC homes) or not satisfied with the way medications were given (2 RAC home).

The top risk factor was "ineffective monitoring process", which occurred in all 18 failed RAC homes. The second frequently stated risk factor was "noncompliance with professional standards and guidelines", occurring in 15 failed homes (83%). Resident's dissatisfaction about the overall medication management was the number three risk factor, indicated in 10 failed RAC homes (56%). Risk factors about staff practice were also quite often mentioned in the reports. These included inconsistency in staff practice to administer medication (8 homes, 44%) and unmonitored staff practice (4 homes, 22%).

Discussion

This study identified 21 risk factors for the RAC homes to fail in medication management in accreditation. It contributes in the knowledge area of aged care risk management.

A medication management process includes prescription, ordering, dispensing, administration, recording and review, storage and disposal [20]. For resident safety, the whole medication management process requires monitoring. We found that "ineffective monitoring process" is the biggest risk factor for the RAC homes to fail in medication management. Our finding is in line with that of another study which found that 70% of adverse drug events in RAC homes were caused by inadequate monitoring of the medication management process [21].

The second major risk factor in medication management is "noncompliance" with professional standards and guidelines. Another notable finding is that staff practice at each stage of medication management is important for medication safety. For example, checking the package of medication will ensure the expired medication would not be administered to a resident. Therefore, developing an effective monitoring system is of critical importance for safe medication management in the RAC homes.

The limitation of this study was the nature of any secondary study, with all the findings drawn from analysis of the accreditation reports. As only 18 out of 3,607 ACAR reports reported failure in medication management for a RAC home, evidence collected from this information source might be limited. In addition, these publicly available reports only summarised high level information. There was inadequate details about what exactly led to failure in medication management. Therefore, we cannot infer how risks happened and what can be done to prevent and control these risks.

The strength of the study is, for the first time, providing a nationwide overview of the reasons for the Australian RAC homes to fail the accreditation outcome in medication management. It gives some insight into aspects of safety-related issues in RAC homes, which may be helpful for future studies.

Conclusion

Using text data mining method, this study identified 21 risk factors for the RAC homes in Australia to fail in medication management accreditation standards. These risk factors fell into 10 categories: overall medication management, medication assessment, ordering, dispensing, stock, administration, monitoring, incident report, staff satisfaction and resident satisfaction. It provides accurate information about which aspects of medication management that the RAC homes in Australia failed in the four year period from 2011 to 2015. This information is useful for RAC home managers and the aged care quality improvement agency to use to improve medication management in this care setting. Future research can be focused on how to best implement a monitoring mechanism to improve medication safety and resident satisfaction with the medication management process.

References

- [1] B.K. Armstrong, J.A. Gillespie, S.R. Leeder, G.L. Rubin, and L.M. Russell, Challenges in health and health care for Australia, *The Medical Journal of Australia* **187** (2007), 485.
- [2] N. Wang, P. Yu, and D. Hailey, Description and comparison of quality of electronic versus paper-based resident admission forms in Australian aged care facilities, *International Journal of Medical Informatics* **82** (2013), 313-324.
- [3] K. Malloch, The electronic health record: an essential tool for advancing patient safety, *Nursing outlook* **55** (2007), 159-161.
- [4] Aged Care Standards and Accreditation Agency, Accreditation overview, 2013.
- [5] Aged Care Standards and Accreditation Agency, Accreditation Standards, 2013.
- [6] V. Adams and R. Sharp, Reciprocity in caring labor: nurses' work in residential aged care in Australia, *Feminist economics* **19** (2013), 100-121.
- [7] Australian Aged Care Quality Agency, Failure to meet the Accreditation Standards, in, 2017.
- [8] N.M. Wilson, L.M. March, P.N. Sambrook, and S.N. Hilmer, Medication safety in residential aged-care facilities: a perspective, *Therapeutic Advances in Drug Safety* **1** (2010), 11-20.
- [9] M.M. Elseviers, R.R. Vander Stichele, and L. Van Bortel, Drug utilization in Belgian nursing homes: impact of residents' and institutional characteristics, *Pharmacoepidemiology and drug safety* **19** (2010), 1041-1048.
- [10] J.A. Doshi, T. Shaffer, and B.A. Briesacher, National estimates of medication use in nursing homes: findings from the 1997 medicare current beneficiary survey and the 1996 medical expenditure survey, *Journal of the American Geriatrics Society* **53** (2005), 438-443.
- [11] A. Szczepura, D. Wild, and S. Nelson, Medication administration errors for older people in long-term residential care, *BMC Geriatrics* **11** (2011), 82.
- [12] S. Qian, P. Yu, D.M. Hailey, and N. Wang, Factors influencing nursing time spent on administration of medication in an Australian residential aged care home, *Journal of Nursing Management* **24** (2016), 427-434.
- [13] W. Ellis, S. Kaasalainen, P. Baxter, and J. Ploeg, Medication management for nurses working in long-term care, *Canadian Journal of Nursing Research* **44** (2012), 128-149.
- [14] T. Jiang and P. Yu, The impact of electronic health records on client safety in aged care homes, *Studies In Health Technology And Informatics* **201** (2014), 116-123.
- [15] J. Scott-Cawiezell, G.A. Pepper, R.W. Madsen, G. Petroski, A. Vogelsmeier, and D. Zellmer, Nursing home error and level of staff credentials, *Clinical Nursing Research* **16** (2007), 72-78.
- [16] H.M. Young, S.L. Gray, W.C. McCormick, S.K. Sikma, S. Reinhard, L. Johnson Trippett, C. Christlieb, and T. Allen, Types, prevalence, and potential clinical significance of medication administration errors in assisted living, *Journal of the American Geriatrics Society* **56** (2008), 1199-1205.
- [17] B.S. Cassidy, The regulatory influence on medication administration in long-term care facilities, *Geriatric Nursing* **26** (2005), 212-215.
- [18] A. Vogelsmeier, J. Scott-Cawiezell, and D. Zellmer, Barriers to safe medication administration in the nursing home: exploring staff perceptions and concerns about the medication use process, *Journal of gerontological nursing* **33** (2007), 5-12.
- [19] Australian Aged Care Quality Agency, Aged care accreditation reports, 2015.
- [20] Department of Health and Ageing, Guiding principles for medication management in residential aged care facilities, Australia Government Department of Health and Ageing, 2012.
- [21] J.H. Gurwitz, T.S. Field, J. Avorn, D. McCormick, S. Jain, M. Eckler, M. Benser, A.C. Edmondson, and D.W. Bates, Incidence and preventability of adverse drug events in nursing homes, *American Journal of Medicine, The* **109** (2000), 87-94.

Address for correspondence

Corresponding author: Associate Professor Ping Yu, Centre for IT-enabled Transformation, School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia. Tel.: +61 242215412; fax: +61 242214045.
E-mail address: ping@uow.edu.au

An Ontology-Based Approach to Estimate the Frequency of Rare Diseases in Narrative-Text Radiology Reports

Charles E. Kahn, Jr.

Department of Radiology and Institute for Biomedical Informatics,
University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

This study sought to use ontology-based knowledge to identify patients with rare diseases and to estimate the frequency of those diseases in a large database of radiology reports. Natural language processing methods were applied to 12,377,743 narrative-text radiology reports of 7,803,811 patients at an academic health system. Using knowledge from the Orphanet Rare Disease Ontology and Radiology Gamuts Ontology, 1,154 of 6,794 rare diseases (17.0%) were observed in a total of 237,840 patients (3.05%). Ninety of 2,129 diseases (4%) with known prevalence less than 1 per 1,000,000 were observed in the database, whereas 100 of 173 diseases (58%) with prevalence greater than 1 per 10,000 were observed; the difference was statistically significant ($p < .00001$). Automated ontology-based search of radiology reports can estimate the frequency of rare diseases, and those diseases with higher known prevalence were significantly more likely to appear in radiology reports.

Keywords:

Information Storage and Retrieval; Knowledge Bases; Rare Diseases

Introduction

A rare disease, sometimes called an "orphan disease," is defined in the United States as a condition that affects fewer than 200,000 Americans (prevalence $< 0.06\%$) and in Europe as one that affects fewer than 1 in 2,000 individuals (prevalence $< 0.05\%$) [1; 8]. Approximately 6,800 rare diseases have been identified [17]. Although each disorder is itself rare – some with prevalence of less than 1 per million – in total, they affect about 5% of the population, which is about one-half the prevalence of diabetes mellitus. People who suffer from rare disease pose substantial clinical and economic burdens [9]. To address these challenges, several national rare-disease registries have been established [23; 24].

The multilingual Orphanet database (<http://www.orpha.net>) provides knowledge about rare diseases for patients, family members, clinicians, and researchers [2; 19]. The database has been derived from the biomedical literature and validated by international experts. Orphanet's database includes the diseases' names, estimated prevalence, mode of inheritance, age of onset, and age of death. The Orphanet Rare Disease Ontology (ORDO) allows computational analysis by providing a standardized vocabulary for rare diseases that expresses relationships between diseases, genes, and other relevant features. ORDO integrates a classification scheme, gene-disease relationships, epidemiological data, and connections with other terminologies. The Radiology Gamuts

Ontology (RGO; <http://www.gamuts.net>), an ontology that links diseases and conditions to their imaging findings, has been integrated with ORDO [4; 12].

The purpose of this study was twofold: (1) to evaluate the large-scale use of natural language processing (NLP) techniques and ontology-based knowledge to identify rare diseases in a large database of radiology reports and (2) to compare the estimated frequencies of the diseases with their known prevalence in the population.

Methods

Radiology Reports Database

This research project was approved by the organization's Institutional Review Board, and was compliant with the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA). A database of reports of 12,377,743 diagnostic and interventional radiology procedures on 7,803,811 distinct patients served as text corpus for the investigation. The reports were generated from September 1988 through October 2015 by more than 350 radiologists in training and attending radiologists at three urban teaching hospitals of a single, U.S. university-based health system. The database included the full text of the radiology reports plus identifying information such as each patient's medical record number, age, and sex.

Ontologies

ORDO (version 1.0.20, modified 14 December 2012) contained 6,794 disease entries. For each disease, a query string was generated. Forty-three Orphanet disease names that contained non-English text characters were converted in automated fashion using UTF-8-to-ASCII mappings. For example, "Behçet disease" was converted to "Behcet disease" and "Waldenström macroglobulinemia" was converted to "Waldenstrom macroglobulinemia." To augment the search, synonyms were obtained using a web-service interface to the Radiology Gamuts Ontology (RGO), a knowledge resource of diseases and their imaging manifestations that has been mapped to ORDO [12]. Because disease names in ORDO and RGO are normalized, possessive forms of disease names were generated as needed. For example, the search for ORDO entity "Waldenström macroglobulinemia" (Orpha ID 33226) included synonyms "Waldenstrom macroglobulinemia", "Waldenstrom's macroglobulinemia", "lymphoplasmacytic lymphoma", "Waldenstrom disease", and "Waldenstrom's disease" (<https://api.gamuts.net/names/742>). The query string specified a search for the appearance of one or all of those terms.

The report database was queried using software located on a server within the health system's secure internal network. The query software automatically presented web-service requests to the application programming interface (API) of a radiology report text search engine (Montage version 3.0.1, Montage Healthcare Solutions, Inc., Philadelphia, PA). The search engine incorporated a proprietary negative-expression (negex) filter to maximize the likelihood of retrieving true-positive cases. The web service used Representational State Transfer (REST) syntax, and returned data encoded in the JavaScript Object Notation (JSON) syntax.

Case Retrieval

Reports were aggregated by patient, and the total number of patients for each disease was recorded. The number of total disease occurrences was computed as the sum of the number of patients with each disease. This sum is the upper bound on the total number of patients, as an individual patient may have had more than one disease. The probability of having at least one disease – that is, the overall prevalence of having a rare disease – was estimated as the complement of the probability of having none of the diseases. Thus, given the frequency of each disease, p_i , one can estimate the probability of having any disease as $p = 1 - \prod_i (1 - p_i)$.

A "disease occurrence" was defined as the presence of a positive mention of a disease in one or more radiology reports of a specific patient. Thus, the number of occurrences for a specific disease represented the number of patients with that disease. "Disease frequency" was defined as the number of patients with a specific disease divided by the total number of patients.

We estimated the case-retrieval system's precision and recall using a sampling technique described previously [10; 13; 16]. The 6,794 diseases and 12,377,743 reports yielded more than 84 billion pairwise combinations, which precluded exhaustive review. We used both "microaveraging" (which considers all disease-report pairs as a single group) and "macroaveraging" (which computes the metrics by disease) of the precision and recall estimates; macroaveraging is preferred because it weights queries equally.

To establish a reference standard, the author reviewed pairs of diseases and reports to indicate whether a positive mention of the disease was present or absent in the report. Negated expressions (e.g., "No evidence of ...") and conjectural expressions ("Rule out ...") were assigned as "absent." To eliminate potential bias, the disease-report pairs were presented in random order, and the reviewer was not aware of whether the disease was meant to be present or absent in the report.

Precision is analogous to positive predictive value; it indicates how many retrieved reports are truly positive. We estimated precision in a sample of 10 positive reports for each of the 20 most common diseases. Reports were reviewed manually to determine if a positive mention of the disease was present (true positive [TP]) or absent (false positive [FP]). Precision was estimated as $TP / (TP + FP)$ for each disease.

Recall is analogous to sensitivity; it measures the fraction of relevant reports that labeled as positive. For the 20 most common diseases, the number of true positive (TP) reports was estimated as the number found with the disease multiplied by the overall precision value. Then, for each disease, we sampled 10 reports in which positive mention of the disease was absent. Those reports should be negative; the "Sample TN" is the number of true negative (TN) reports among the 10 sampled for each disease. Based on the Sample TN value, we

extrapolated to the entire set of negative reports. Recall was computed as $TP / (TP + TN)$. The example provided in [13] helps explain the calculation. For values of precision, P, and recall, R, we computed the F1 score, the harmonic mean of precision and recall, as $2 \cdot P \cdot R / (P + R)$.

Results

Females constituted 4,694,083 patients (60.2%) and were the subject of 7,234,080 reports (58.4%). Patients under 20 years of age comprised 489,974 reports (3.9%) (Figure 1). Of the 6,794 Orphanet rare diseases, 1,154 (17.0%) were observed in at least one patient in the report database. There were a total of 237,840 disease occurrences in the report database, which corresponded to 3.05% of all patients. The aggregate prevalence of rare disease was estimated at 3.00%, or approximately 234,324 patients.

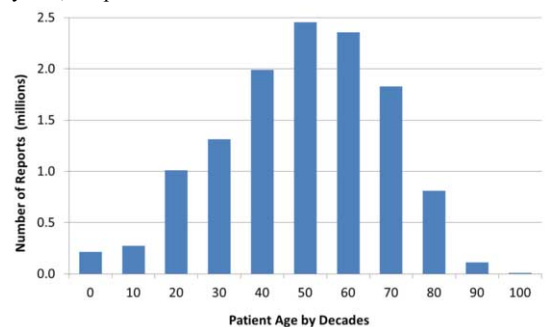


Figure 1— Age distribution histogram.

The 20 rare diseases that appeared most frequently in the radiology reports accounted for 103,402 (43.4%) of the 237,840 disease occurrences (Table 1). Among the least frequently appearing diseases, 234 appeared in only 1 patient, and 368 diseases appeared in 2 to 10 patients. A sample of 20 diseases that did not appear in the collection of reports is shown in Table 2.

Table 1— The 20 most frequent rare diseases in the radiology report database.

Disease Name	No. of Patients	Prevalence (per million)	
		Observed	OrphaNet
Tuberculosis	16,698	2,140	100–500
Sarcoidosis	12,432	1,593	100–500
Meningioma	12,340	1,581	10–90
Adult acute respiratory distress syndrome	8,845	1,133	100–500
Hepatocellular carcinoma	7,783	997	100–500
Arachnoid cyst	5,426	695	-
Idiopathic pulmonary fibrosis	5,352	686	100–500
Multiple myeloma	4,945	634	100–500
Benign schwannoma	3,715	476	10–90
Lyme disease	3,313	425	-
Hypersensitivity pneumonitis	3,172	406	-
Periventricular leukomalacia	2,949	378	-
Pancreatic carcinoma	2,917	374	100–500
Holt-Oram syndrome	2,618	335	-
Aspergillosis	2,320	297	-
Esophageal carcinoma	2,153	276	10–90
Hodgkin lymphoma, classical	2,088	268	100–500
Arachnoiditis	1,988	255	-
Small cell lung cancer	1,190	152	100–500
Tetralogy of Fallot	1,158	148	100–500
TOTAL	103,402		

Table 2 – Random sample of 20 rare diseases for which no patients were found in the radiology report database.

Disease Name

- pachygyria – intellectual deficit – epilepsy
- complete atrioventricular canal – Fallot tetralogy
- multiple epiphyseal dysplasia type 1
- Marshall syndrome
- Keipert syndrome
- Chudley-Lowry syndrome
- Böök syndrome
- dysostosis Stanescu type
- xeroderma pigmentosum – Cockayne syndrome complex
- acrofacial dysostosis Palagonia type
- arterial dissection – lentiginosis
- ossification anomalies – psychomotor development delay
- Potocki-Shaffer syndrome
- craniosynostosis – synostoses – hypertensive nephropathy
- lethal multiple pterygium syndrome
- spina bifida – hypospadias
- trismus-pseudocamptodactyly syndrome
- COFS syndrome
- auriculoosteodysplasia
- segmental outgrowth – lipomatosis – arteriovenous malformation – epidermal nevus

Based on review of 10 positive and 10 negative reports for each of the 20 most frequent diseases (Table 3), estimated information-retrieval metrics were: precision = 0.955, recall = 0.900, and F1 = 0.927.

Table 3 – Estimation of precision and recall using random samples of 10 positive and 10 negative cases for the 20 most common diseases in the database. TP = true positive, TN = true negative.

Disease	N	Sample		Precision	Recall
		TN	TP		
Acute respiratory distress syndrome	8845	10	10	1.0000	1.0000
Arachnoid cyst	5426	10	10	1.0000	1.0000
Arachnoiditis	1988	10	9	0.9000	1.0000
Aspergillosis	2320	10	10	1.0000	1.0000
Benign schwannoma	3715	10	10	1.0000	1.0000
Esophageal carcinoma	2153	10	10	1.0000	1.0000
Hepatocellular carcinoma	7783	10	9	0.9000	1.0000
Hodgkin lymphoma, classical	2088	9	8	0.8000	0.0003
Holt-Oram syndrome	2618	10	10	1.0000	1.0000
Hypersensitivity pneumonitis	3172	10	10	1.0000	1.0000
Idiopathic pulmonary fibrosis	5352	10	10	1.0000	1.0000
Lyme disease	3313	10	9	0.9000	1.0000
Meningioma	12340	10	10	1.0000	1.0000
Multiple myeloma	4945	10	10	1.0000	1.0000
Pancreatic carcinoma	2917	10	10	1.0000	1.0000
Periventricular leukomalacia	2949	10	7	0.7000	1.0000
Sarcoidosis	12432	10	10	1.0000	1.0000
Small cell lung cancer	1190	9	9	0.9000	0.0001
Tetralogy of Fallot	1158	10	10	1.0000	1.0000
Tuberculosis	16698	10	10	1.0000	1.0000
				0.9550	0.9000

The rare diseases with higher prevalence in ORDO occurred more frequently in the report database (Table 4): 90 of 2,129 diseases (4.2%) with prevalence less than 1 per 1,000,000 were observed in the database, whereas 100 of 173 diseases (57.8%) with prevalence greater than 1 per 10,000 were observed; chi-square test showed a statistically significant difference ($p < .00001$). Of the 2,287 ORDO terms that had matching RGO terms, 740 diseases (32.4%) appeared in the clinical reports; 414 of 4,210 diseases (9.8%) not matched to

RGO appeared in reports; the difference was statistically significant ($p < .00001$).

Table 4 – Frequency of diseases grouped by prevalence as defined in Orphanet. For each group, the table indicates the number of diseases in that group (and percentage of all Orphanet diseases), the number of diseases observed in the report database, and the number of patients. Prevalence listed as "Unknown" includes those in Orphanet as "unknown" or "unspecified." The group with prevalence greater than or equal to 100 per million includes the Orphanet prevalence categories "1-5 / 10 000," "6-9 / 10 000," and "> 1 / 1000." The mean observed prevalence is the mean number of patients per disease in each group, divided by the study's total number of patients.

Orphanet Prevalence (per million)	No. of Diseases (%)		No. of Diseases Observed (%)		No. of Patients	Mean Observed Prevalence (per million)
Unknown	3,926	58%	705	18%	85,666	2.80
< 1	2,139	31%	90	4%	1,834	0.11
1 - 9	207	3%	79	38%	1,898	1.17
10 - 99	349	5%	180	52%	48,237	17.71
≥ 100	173	3%	100	58%	100,205	74.22
Total	6,794	100%	1,154	17%	237,840	4.49

Discussion

Rare diseases present important medical and social challenges [22]. The current study demonstrates the ability to apply ontology-based knowledge to estimate the frequency of almost 6,800 rare diseases in a large corpus of narrative-text radiology reports. The frequency of diseases in the report database showed overall correspondence with known prevalence data, in that diseases with unknown or smaller prevalence values (e.g., < 1 per million) were less likely to appear in the report database. Similarly, rare diseases in Orphanet that had been mapped to diseases in RGO were significantly more likely to appear in the set of reports, probably because the diseases were more common and because they were more likely to involve radiological findings. Frequencies for individual disorders did not necessarily correspond closely to known prevalence values, but the observed values could be useful to rank diseases for diagnosis based on imaging findings. As such, one might consider them to represent "radiological prevalence" values, in that they indicate the frequency with which the conditions appear in radiology reports.

The estimated disease frequencies may have been subject to several biases. First, the study used radiology reports from a single academic medical institution, which may have reflected the nature of a university-based specialty clinical practice and/or the conditions for which the institution had particular expertise. Second, the report database had a limited number of children; many rare diseases have onset and mortality in childhood, and such patients were underrepresented in the patient population. An affiliated children's hospital, whose reports were not included, likely attracted pediatric rare-disease patients preferentially. Third, many rare diseases do not entail imaging findings, and may not have been mentioned in the reports. Fourth, radiology reports do not necessarily contain comprehensive information about the patient: the referring provider may not have mentioned the disease in the clinical information, and the radiologist may not have reported it.

Adjustment of several of the search queries was needed in order to accommodate idiosyncratic aspects of the Montage

search engine and terms in the ontologies. For example, initial queries for "small cell lung cancer" identified reports containing the phrase "non-small cell lung cancer"; similarly, searches for "Hodgkin's lymphoma" resulted in reports containing "non-Hodgkin's lymphoma." After the initial queries were adjusted to explicitly exclude the "non-" prefix, the revised queries resulted in high precision and recall. The Gamuts ontology included the term "TOF" as a synonym for "tetralogy of Fallot"; that abbreviation, however, yielded false-positive disease occurrences where it was detected in a number of reports where "3D TOF" had been used to indicate "three-dimensional time-of-flight." Given its overwhelming use in MRI reports, "TOF" was excluded as a search term.

Systems for natural language processing (NLP) have shown high performance in identifying and classifying radiology reports, and offer a promising approach to extract measurable information from conventional narrative ("free-text") reports [5; 18]. An NLP system successfully determined the presence of more than 20 clinical indications and imaging findings from a database of 889,921 chest radiographic reports [11]. Negative-expression ("negex") filtering techniques have identified positive and negative instances of imaging findings with accuracy in excess of 97% [6; 7]. PRESTO, the predecessor to the commercial system used in the current study, showed 91% precision and 97% recall in detecting adrenal findings in 32,974 abdominal CT reports [25].

To understand the challenges inherent in automated approaches to report classification, it is instructive to consider two diseases with known prevalence greater than 1 in 1,000. The first, ventricular septal defect (VSD; Orpha number 1480), was seen in 844 patients; this observed frequency of 108 patients per million was an order of magnitude less than the prevalence specified in Orphanet. The second disease, "Non-classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency" (Orpha number 95698), was not observed in any report. One can speculate that the disease's longer name made it less likely to be identified in radiology reports. The term "congenital adrenal hyperplasia" was found in 145 patients, and "21-hydroxylase deficiency" was found in only two patients.

Ongoing work centers on the use of ontology-based knowledge and imaging phenotypes, that is, findings identified in imaging studies. ORDO has been linked to the Human Phenotype Ontology (HPO), a resource that categorizes the manifestations of diseases, predominantly congenital or genetic disorders [15; 21]. One research opportunity is to identify constellations of phenotypic findings that may point to patients with otherwise undiagnosed rare diseases. Integration of HPO and the Gamuts ontology is in development to better identify phenotypic manifestations of disease. The integration of these ontologies enables the investigations to identify new patterns of association and/or causality among imaging findings and diseases. Ongoing work includes the analysis of the frequency of a more comprehensive set of diseases and imaging findings to explore their prior and conditional probabilities. This growing effort in "deep phenotyping" integrates disease information and detailed phenotype information, such as imaging biomarkers, to support precision medicine [20].

Conclusion

Text-search and NLP techniques can be applied to estimate the frequency of rare diseases in a large database of radiology reports. Ontologies such as ORDO and RGO help organize

medical knowledge so that automated systems can make inferences from that knowledge [3]. The increasing availability of large datasets in radiology and the tools to analyze them can lead to better data-driven understanding of radiology practice [14].

References

- [1] Public Law 107-280: Rare Diseases Act of 2002, in, U.S. Government Printing Office, Washington, DC, 2002.
- [2] S. Aymé, B. Urbero, D. Oziel, E. Lecouturier, and A.C. Biscarat, Information sur les maladies rares: le projet orphanet [Information on rare diseases: the Orphanet project], *La Revue de médecine interne* **19 Suppl 3** (1998), 376S-377S.
- [3] O. Bodenreider, Biomedical ontologies in action: role in knowledge management, data integration and decision support, *Yearb Med Inform* (2008), 67-79.
- [4] J.J. Budovec, C.A. Lam, and C.E. Kahn, Jr., Radiology Gamuts Ontology: differential diagnosis for the Semantic Web, *RadioGraphics* **34** (2014), 254-264.
- [5] T. Cai, A.A. Giannopoulos, S. Yu, T. Kelil, B. Ripley, K.K. Kumamaru, F.J. Rybicki, and D. Mitsouras, Natural language processing technologies in radiology research and clinical applications, *RadioGraphics* **36** (2016), 176-191.
- [6] P.A. Dang, M.K. Kalra, M.A. Blake, T.J. Schultz, M. Stout, E.F. Halpern, and K.J. Dreyer, Use of Radcube for extraction of finding trends in a large radiology practice, *Journal of Digital Imaging* **22** (2009), 629-640.
- [7] K.J. Dreyer, M.K. Kalra, M.M. Maher, A.M. Hurier, B.A. Asfaw, T. Schultz, E.F. Halpern, and J.H. Thrall, Application of recently developed computer algorithm for automatic classification of unstructured radiology reports: validation study, *Radiology* **234** (2005), 323-329.
- [8] European Organisation for Rare Diseases, Rare Diseases: Understanding this Public Health Priority, in, Paris, France, 2005.
- [9] J.M. Graf von der Schulenburg and M. Frank, Rare is frequent and frequent is costly: rare diseases as a challenge for health care systems, *Eur J Health Econ* **16** (2015), 113-118.
- [10] W. Hersh, Evaluation of biomedical text-mining systems: lessons learned from information retrieval, *Briefings in Bioinformatics* **6** (2005), 344-356.
- [11] G. Hripesak, J.H. Austin, P.O. Alderson, and C. Friedman, Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports, *Radiology* **224** (2002), 157-163.
- [12] C.E. Kahn, Jr., Integrating ontologies of rare diseases and radiological diagnosis, *Journal of the American Medical Informatics Association* **22** (2015), 1164-1168.
- [13] C.E. Kahn, Jr. and D.L. Rubin, Automated semantic indexing of figure captions to improve radiology image retrieval, *J Am Med Inform Assoc* **16** (2009), 380-386.
- [14] A.P. Kansagra, J.P. Yu, A.R. Chatterjee, L. Lenchik, D.S. Chow, A.B. Prater, J. Yeh, A.M. Doshi, C.M. Hawkins, M.E. Heilbrun, S.E. Smith, M. Oselkin, P. Gupta, and S. Ali, Big Data and the future of radiology informatics, *Academic Radiology* **23** (2016), 30-42.
- [15] S. Kohler, S.C. Doelken, C.J. Mungall, S. Bauer, H.V. Firth, I. Bailleul-Forestier, G.C. Black, D.L. Brown, M. Brudno, J. Campbell, D.R. FitzPatrick, J.T. Eppig, A.P. Jackson, K. Freson, M. Girdea, I. Helbig, J.A. Hurst, J. Jahn, L.G. Jackson, A.M. Kelly, D.H. Ledbetter, S. Mansour, C.L. Martin, C. Moss, A. Mumford, W.H. Ouwehand, S.M. Park, E.R. Riggs, R.H. Scott, S. Sisodiya, S. Van Vooren, R.J. Wapner, A.O. Wilkie, C.F. Wright, A.T. Vulto-van Silfhout, N. de Leeuw, B.B. de Vries, N.L. Washington, C.L. Smith, M. Westerfield, P. Schofield, B.J. Ruef, G.V. Gkoutos, M. Haendel, D. Smedley, S.E. Lewis, and P.N. Robinson, The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data, *Nucleic Acids Research* **42** (2014), D966-974.
- [16] D.D. Lewis, Evaluating text categorization, in: *Proceedings of the Speech and Natural Language Workshop*, Morgan Kaufmann, Asilomar, CA, 1991, pp. 312-318.
- [17] National Center for Advancing Translational Sciences, Office of Rare Diseases Research (ORDR), in, National Institutes of Health, Bethesda, MD, 2013.
- [18] E. Pons, L.M.M. Braun, M.G.M. Hunink, and J.A. Kors, Natural Language Processing in radiology: a systematic review, *Radiology* **279** (2016), 329-343.
- [19] A. Rath, A. Olry, F. Dhombres, M.M. Brandt, B. Urbero, and S. Aymé, Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users, *Human Mutation* **33** (2012), 803-808.
- [20] P.N. Robinson, Deep phenotyping for precision medicine, *Human Mutation* **33** (2012), 777-780.

- [21] P.N. Robinson and S. Mundlos, The human phenotype ontology, *Clinical Genetics* **77** (2010), 525-534.
- [22] A. Schieppati, J.I. Henter, E. Daina, and A. Aperia, Why rare diseases are an important medical and social issue, *Lancet* **371** (2008), 2039-2041.
- [23] D. Taruscio, L. Vittozzi, R. Choquet, K. Heimdal, G. Iskov, Y. Kodra, P. Landais, M. Posada, R. Stefanov, C. Steinmueller, E. Swinnen, and H. Van Oyen, National registries of rare diseases in Europe: an overview of the current situation and experiences, *Public Health Genomics* **18** (2015), 20-25.
- [24] A.C. Zoni, M.F. Dominguez Berjon, E. Barcelo, M.D. Esteban Vasallo, I. Abaitua, J. Jimenez Villa, M. Margolles Martins, C. Navarro, M. Posada, J.M. Ramos Aceitero, C. Vazquez Santos, O. Zurriaga Llorens, J. Astray Mochales, and R.D.R.G. Spain, Identifying data sources for a national population-based registry: the experience of the Spanish Rare Diseases Registry, *Public Health* **129** (2015), 271-275.
- [25] J.J. Zopf, J.M. Langer, W.W. Boonn, W. Kim, and H.M. Zafar, Development of automated detection of radiology reports citing adrenal findings, *Journal of Digital Imaging* **25** (2012), 43-49.

Address for correspondence

Charles E. Kahn, Jr., MD, MS, Department of Radiology, University of Pennsylvania Perelman School of Medicine, 3400 Spruce Street, Philadelphia, PA 19104 USA. ckahn (at) upenn.edu

On Incentives for Open Access Publishing: A Survey at IMIA's Annual General Assembly During HEC2016

Stefanie Kuballa^{a,*}, Mareike Schulze^{a,*}, Corinna Mielke^a, Olaf Gefeller^b, Reinhold Haux^a

^a Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig – Institute of Technology and Hannover Medical School, Braunschweig, Germany

^b Institute for Medical Informatics, Biometry and Epidemiology, University of Erlangen-Nürnberg, Erlangen, Germany

* These authors contributed equally to this work

Abstract

Open access provides an alternative opportunity of publishing research articles besides the traditional subscription-based publication model. Facilitating higher visibility, accessibility, and dissemination, among others, open access addresses modern needs of our information and communication society. Though these factors are vital, there also do exist initial problems to become prevalent. Especially, costs like article publication charges seem to have an impact on the author's decision not to publish open access. In order to gain a deeper insight on incentives for open access publishing (as well as on barriers for not publishing open access), we developed a specific survey within the scope of our Trans-O-MIM project. Conducted at HEC2016, we primarily involved participants of the Annual General Assembly of the International Medical Informatics Association (IMIA). As main incentives and motivation for open access publishing had been identified: article indexing, impact factor, accessibility/availability, dissemination, and visibility.

Keywords:

Open Access Publishing, Motivation, Surveys and Questionnaires

Introduction

In recent years, a change in the publication landscape can be observed. Besides the traditional publication model, an alternative model is about to establish itself [1-3]. Open access publishing aims to fulfill needs of modern information and communication society. The Internet allows publishing journal articles or even whole journals online and the open access model takes on this development by offering all content freely available to readers. Also, many research funding organizations and governments have changed their policies and demand to publish funded research outcomes open access [4, 5]. To finance this new publication model, authors are often charged so-called article publication charges (APC).

Nearly all journals offer their authors an open access option [6]. But the mere fact that there is a new possibility of publishing research articles does not necessarily mean that authors make use of it. Currently, only about 13% of all research papers are published open access [7]. Besides all positive attributes of open access publishing - for example higher visibility and citation rates - there are also many concerns. Significant barriers to open access publishing are an often-higher effort for the authors and the above-mentioned publication charges [8, 9].

When authors make a decision how to publish their work - traditional or open access - they are usually influenced by certain factors. Within the framework of the project "Trans-O-MIM" (full title "Strategies, models and evaluation metrics for the goal-oriented, stepwise, sustainable and fair transformation of established subscription-based scientific journals into open-access-based journals with *Methods of Information in Medicine* as example") [10, 11], funded by the German Research Foundation (DFG), we intended to identify incentives tempting authors into publishing open access. These results are conceived to help transform the journal *Methods of Information in Medicine* successfully into an open access journal.

The goal is to explore factors that influence authors positively in matters of open access publishing. To compose a good, and later on accepted, open access business model, it is crucial to understand the needs and wishes of researchers and authors, respectively. By means of a survey among medical informatics scientists we explored decisive incentives.

Most importantly in the decision-making process for a journal is Impact Factor, a fast turnaround time and most of all the publication charges. This leads to the question of to which limit these charges are considered as appropriate and tolerable. In this context, it is moreover presumed that most of the authors do not have that much experience in open access publishing. For the utilization of the survey results this hypothesis shall be investigated additionally.

At the beginning of this investigation several general research questions were collected. These were broadly diversified in terms of incentives for open access publishing. General thoughts in this context were (e.g.):

- how experienced authors are in open access publishing
- if incentives are actually required and if so who needs to be incentivized
- whether there are other factors that may influence authors in their decision for publication

This constituted the initial point for more detailed questions derived from a broader context and the methodology of surveys and survey design.

In the subsequent sections, we describe the questionnaire intended to get answers to these questions, its conduct and results. Furthermore, we will discuss these findings pertaining to an adequate open access business model and close with a conclusion.

Methods

We designed a questionnaire in order to ask experts from the fields of medical informatics, biometry, and epidemiology about their knowledge regarding existing open access journals and about their incentives to publish research articles open access. The paper-based version of our questionnaire was created in cooperation with and typeset by Schattauer Publishers (Figure 1); each optimized in terms of presentiveness and usability.

Which incentives tempt authors into publishing Open Access?

Dear participants of HEC 2016,

please allow us some questions in the context of the DFG-project "Trans-O-MIM" (www.pli.de/en/forschung/impulse/trans-o-mim.html), which deals with the transition of subscription-based scientific journals into open-access-based journals with "Methods of Information in Medicine" as example.

Please fill out one form, either on paper or electronically (see QR-Code). Thank you!

As a reward for your help we will offer a free Open Access article in "Methods of Information in Medicine" amongst all participants. If you want to take part please fill in your e-mail address:

1. Which factors influence your decision to choose a journal? (importance on a scale from 1 = not important to 10 = very important; 0 = not applicable)

- Articles are indexed e.g. in PubMed Central
- Impact Factor of a journal
- Publication charges
- Prestige of a journal
- Visibility and dissemination
- Topic fits optimally to the journal
- Ability to attach supplementary material/data
- Members of the editorial board
- A fast turnaround time
- Anything else, please specify

2. Have you ever published a journal manuscript in Open Access?

Yes ☐ No ☐ n.a.

3. Do you know any Open Access journals in the field of Medical Informatics, Biometry and / or Epidemiology?

Yes ☐ No ☐ n.a.

4. If you answered the previous question with "yes", have you already published in one of them?

Yes ☐ No ☐ n.a.

5. What is your motivation to publish in an Open Access journal? (in keywords)

-
-
-

6. Which amount of article processing charges (APC) is appropriate and which is tolerable in your opinion?

	appropriate	tolerable	n.a.
1) up to 1.000 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2) 1.001 € to 1.500 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3) 1.501 € to 2.000 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4) 2.001 € to 2.500 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5) more than 2.501 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

7. What are the three main reasons for you NOT to publish your work in an open access journal? (in keywords)

-
-
-

8. Do you have any other comments, e.g. on the incentives to publish in an Open Access journal?

www.umfrage.de

Figure 1 – paper-based version of survey

Our questionnaire contains a short introduction about the "Trans-O-MIM" DFG-project and a guidance how to fill in the questionnaire. The questionnaire itself consists of eight questions, namely:

- Which factors influence your decision to choose a journal? (importance on a scale from 1 = not important to 10 = very important; 0 = not applicable)
 - Articles are indexed e.g. in PubMed Central
 - Impact Factor of a journal
 - Publication charges
 - Prestige of a journal
 - Visibility and dissemination
 - Topic fits optimally to the journal
 - Ability to attach supplementary material/data
 - Members of the editorial board
 - A fast turnaround time
 - Anything else, please specify
- Have you ever published a journal manuscript in open access?
- Do you know any open access journals in field of Medical Informatics, Biometry and / or Epidemiology?

- If you answered the previous question with "yes", have you already published in one of them?
- What is your motivation to publish in an open access journal?
- Which amount of article processing charges (APC) is appropriate and which is tolerable in your opinion?
 - Up to 1.000 €
 - 1.001 € to 1.500 €
 - 1.501 € to 2.000 €
 - 2.001 € to 2.500 €
 - More than 2.501 €
- What are the three main reasons for you NOT to publish your work in an open access journal?
- Do you have any other comments, e.g. on the incentives to publish in an open access journal?

Before our questionnaire was printed, a review with three experts in the fields of publishing and medical informatics was done. The experts mentioned were the current and the senior Editor-in-Chief of the journal "Methods of Information in Medicine" and a member of the DFG-committee for research libraries and information systems. All questions were discussed and improved based on the input of these experts. Afterwards, an online version of our survey was developed using a specific tool called "eSurvey Creator" (<https://www.esurveycreator.com>) (as shown in Figure 2).

Which Incentives Tempt Authors Into Publishing Open Access?

Which amount of article processing charges (APC) is appropriate and which is tolerable in your opinion?

	appropriate	tolerable	n.a.
up to 1.000 €	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.001 € to 1.500 €	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
1.501 € to 2.000 €	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2.001 € to 2.500 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
more than 2.501 €	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Next Done

Figure 2 – online version of survey

In order to reach a large group of experts we decided to place our survey at HEC (Health Exploring Complexity: An Interdisciplinary Systems Approach; GMDS & DGEpi & IEA-EEF annual meeting, Medical Informatics Europe – MIE2016; 28 August – 2 September 2016), one of the main international conferences in the field of medical informatics, biometry and epidemiology. HEC 2016 had been visited by about 1,400 conference participants. During this conference, some special meetings of expert associations were held; e.g. IMIA's Annual General Assembly among others. We used this meeting to directly contact and inform the participants of

IMIA's Annual General Assembly about our survey and invited all of them to participate. In addition, we also offered all conference participants to take part in our survey.

In order to merge survey results, we digitalised the completed paper-based questionnaires before evaluation. Responses presented in results section are given in original. Only free text answers were summarized and numerated for ease of exposition.

Results

35 [70%] out of 50 participants of IMIA's 2016 General Assembly took part in our survey. Furthermore, 27 additional participants of HEC2016 filled-in the survey. Results shown in this section are based upon questionnaires gathered from IMIA's 2016 General Assembly.

Question 1

Starting with the first question dealing with influence factors on journal decisions, the following results were achieved.

Table 1 – "Which factors influence your decision to choose a journal" (question 1) – results in absolute values ^(*)

	10	9	8	7	6	5	4	3	2	1	0
a.	18	8	3	1	3	-	-	-	-	-	2
b.	12	5	8	2	2	4	-	1	-	-	1
c.	8	1	7	6	3	6	-	2	-	1	1
d.	6	6	10	6	2	-	-	1	1	-	3
e.	8	5	11	4	2	1	1	1	1	1	-
f.	4	5	11	3	5	5	-	-	1	1	-
g.	1	-	2	5	4	5	6	3	4	3	2
h.	-	2	5	4	3	6	2	1	6	5	1
i.	-	2	6	9	3	3	3	1	-	3	1
j.	*	*	*	*	*	*	*	*	*	*	*

^(*) Additional influence factors specified by participants were: "Appear on the government list of journals that release budget for the university", "brings out future updates on the topics published", "good and transparent review process (with feedback)", "good experience, excellent review process, well written constructive reviews", "government policy", "opened review process", "Quality of reviews" and "submission system".

Questions 2

Going on, 26 [74.3%] from our 35 participants have already published a journal manuscript in open access, whereas 8 [22.9%] did not while 1 [2.9%] stated "n.a".

Questions 3

26 participants answered the question if they know any open access journals in the field of medical informatics, biometry and / or epidemiology. 17 [65.4%] out of these 26 participants knew at least one open access journal whereas 4 [15.4%] did not, while 5 [19.2%] stated "n.a.". Journals mentioned more than once were "Journal of Medical Internet Research (JMIR)", "BioMed Central (BMC)" and "Public Library of Science (PLOS)".

Questions 4

Those 17 participants who answered 'yes' to the previous question were asked if they have already published in one of the journals mentioned. 8 [50%] did whereas 7 [43.8%] did not, while 1 [6.3%] stated "n.a.". Journals mentioned more than once here were BMC and JMIR. One participant did not answer this question.

Question 5

Regarding the motivation to publish in an open access journal - answered by 33 participants - the following answers were given more than once:

1. accessibility / availability – 15 times [45.46%]
2. dissemination – 9 times [27.27%]
3. visibility – 8 times [24.24%]
4. citations / fast process / (funding) policy – 4 times each [12.12% each]
5. fast review – 2 times [6.06%]

Question 6

Talking about article publication charges (APCs) all 35 participants answered the question with the following results.

Table 2 – "Which amount of article processing charges (APC) is appropriate and which is tolerable in your opinion" (question 6) – results in absolute values

	appropriate	tolerable	n.a.
up to 1,000€	12 [34%]	12 [34%]	11 [31%]
1,001€ to 1,500€	6 [17%]	9 [26%]	20 [57%]
1,501€ to 2,000€	2 [6%]	10 [29%]	23 [66%]
2,001€ to 2,500€	1 [3%]	1 [3%]	33 [94%]
more than 2,501€	1 [3%]	1 [3%]	33 [94%]

When no answer was given on paper-based version here, we classified the answer as "n.a".

Question 7

When being asked about the main reasons for them not to publish their work in an open access journal - answered by 31 participants - the following answers were given more than twice:

1. costs – 21 times [68%]
2. reputation – 6 times [19%]
3. no funding / review (process) – 4 times each [13% each]
4. impact factor / quality – 3 times each [10% each]

Question 8

With respect to the last question, two out of six comments gave additional information addressing open access and were not already covered by another question of the survey:

- "There should be a means to waive for discount. The page charges for the author depending on the quality of content, value addition to the confirmity and economic status of the authors(s)."
- "Having so many invitations to become a reviewer / editorial board from open access journals. I doubt the quality of reviewers. I hope MIM make full use of current selected reviewers."

Discussion

Even though 60 percent of our participants already published in open access journals, many of them seem less ambitious in addressing the open access paradigm in depth. When talking about open access, costs will always be one of the first remarks on the table.

Regarding explanatory factors for journal decision in general, article indexing has been declared as most important; and more important than the impact factor and/or prestige of a journal as second priority. Visibility and dissemination as

third priority had also been declared as being more important than publication charges.

Among the main five motivations for publishing open access, accessibility and availability, dissemination, visibility, citations and a fast processing time had been mentioned.

On the downside, participants had been concerned about costs (above all), reputation, no funding, the review process, impact factor and quality.

In terms of costs, article publication charges (APCs) up to 1,000€ had been thought of as being appropriate whereas article publication charges (APCs) up to 2,000€ had been thought of as being tolerable in general tendency.

Limitations

A limitation to this survey is the small number of participants. For the evaluation of our questions we only had 35 answered questionnaires. However, these participants were experts in their field (medical informatics, biometry and epidemiology) so their opinions may have a particularly high explanatory power. Moreover, this survey was kept rather general and preset leading solely to first insights.

Future perspectives

Based on the presented survey and its results we intend to conduct another significantly larger study. Therein researchers in the field of medical informatics will be asked about incentives for open access publishing in detail. The study may also be accompanied by semi-structured interviews of experts to gain new ideas in terms of incentivizing open access publishing for authors.

Conclusion

In conclusion, there have not been any novel findings at all based on the survey. There is indication that open access does not seem to be a big talking point among scientific researchers according to our findings. If so, what is the basis for this circumstance? Although all influence factors seem to be (and have already been) well-known, it turned out to be more difficult than expected to successfully transform well-established subscription-based scientific journals into open access journals.

Although number of participants was small we wanted to share our findings as they provide indicator first impressions. Based on these insights, a follow-up study as mentioned above is going to take place in 2017/2018.

Acknowledgements

This research has been supported by Deutsche Forschungsgemeinschaft under grant HA 1438/17-1.

References

- [1] M. Laakso, P. Welling, H. Bukvova, L. Nyman, B.C. Björk, and T. Hedlund, The development of open access journal publishing from 1993 to 2009, *PLoS One* **6** (2011), e20961.
- [2] K. Kurata, T. Morioka, K. Yokoi, and M. Matsubayashi, Remarkable growth of open access in the biomedical field: analysis of PubMed articles from 2006 to 2010, *PLoS One* **8** (2013), e60925.
- [3] A.J. Wolpert, For the sake of inquiry and knowledge – the inevitability of open access, *N Engl J Med* **368** (2013), 785-787.
- [4] Knowledge Exchange. Bristol: Knowledge Exchanged. open access. [cited 2016 Oct 27] Available from: <http://www.knowledge-exchange.info>.
- [5] European Commission, *open innovation, open science, open to the world – a vision for Europe*, European Commission, Brussels, 2016. [cited 2016 Oct 27] Available from: <http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKl0416263/?CatalogCategoryID=Gj0KABst5F4AAAEjsZAY4e5L>.

- [6] S. Kuballa, Open Access Publishing in the Field of Medical Informatics, *J Med Syst* **41** (2017), 82.
- [7] R. Schimmer, K.K. Geschuhn, and A. Vogler, *Disrupting the subscription journals' business model for the necessary large-scale transformation to open access*, 2015. [cited 2016 Oct 27] Available from: <http://hdl.handle.net/11858/00-001M-0000-0026-C274-7>.
- [8] C. Haug, The downside of open-access publishing, *N Engl J Med* **368** (2013), 791-793.
- [9] R. van Noorden, open access: The true cost of science publishing, *Nature* **495** (2013), 426-429. Errata in **496**, 151 and **499**, 19.
- [10] R. Haux, S. Kuballa, M. Schulze, C. Böhm, O. Gefeller, J. Haaf, P. Henning, C. Mielke, F. Niggemann, A. Schürg, and D. Bergemann, Exploring possibilities for transforming established subscription-based scientific journals into open access journals. Present situation, transformation criteria, and exemplary implementation within Trans-O-MIM, *Methods Inf Med* **55** (2016), 481-487.
- [11] S. Koch, R. Haux, O. Gefeller, I.N. Sarkar and D. Bergemann, Methods Open – A New Journal Track Starting in 2017, *Methods Inf Med* **55** (2016), 478-480.

Addresses for correspondence

Stefanie Kuballa and Mareike Schulze
 Peter L. Reichertz Institute for Medical Informatics
 University of Braunschweig - Institute of Technology and Hannover
 Medical School
 Muehlenpfordtstr. 23
 38106 Braunschweig
 Germany

E-mail: {stefanie.kuballa, mareike.schulze}@plri.de

Peer Review of Clinical Information Models: A Web 2.0 Crowdsourced Approach

Heather Leslie^{a,b}, Silje Ljosland Bakke^{a,c}

^a Co-lead, Clinical Model program, openEHR Foundation,

^b CMIO, Ocean Health Systems, Australia

^c Information Architect, Nasjonal IKT HF, Norway

Abstract

Over the past 8 years the openEHR Clinical Model program has been developing a Web 2.0 approach and tooling to support the development, review and governance of atomic clinical information models, known as archetypes. This paper describes the background and review process, and provides a practical example where cross standards organisation collaboration resulted in jointly agreed clinical content which was subsequently represented in different implementation formalisms that were effectively semantically aligned. The discussion and conclusions highlight some of the socio-technical benefits and challenges facing organisations who seek to govern atomic clinical information models in a global and collaborative online community.

Keywords:

Informatics; Crowdsourcing; Common Data Elements

Introduction

Historically, most collaboration in the health technology domain has been through formal balloting of message or document specification standards within standards development organisations (SDOs) such as ISO TC215 or HL7 International. This approval process has had some significant success over many years in supporting interoperability of health data, however this approach is not transparent, responsive or agile enough for development, maintenance and governance of larger numbers of more atomic clinical information models.

The Clinical Models Program [1] at the openEHR Foundation [2] has developed an alternative, crowdsourced approach to the development, publication and governance of the openEHR clinical information models, known as archetypes. This methodology emphasises openness, transparency and accountability to the community.

The Clinical Knowledge Manager tool was developed directly as a result of the experience of openEHR Clinical Program leadership in working with distributed groups using technical tools such as widely used software versioning and revision control systems. It quickly became apparent that there was not only a need for versioning governance but also life cycle and naming governance plus a critical process to ensure appropriate review of the archetypes to ensure that each was fit for use in implementations.

This openEHR methodology is now in use by a number of national and jurisdictional eHealth programs around the world who are using archetypes published in this manner to underpin their local health IT infrastructure.

Background

The Clinical Model program at the openEHR Foundation is responsible for management of a set of clinical information models, known as archetypes, on behalf of the international openEHR community. The scope of responsibility includes:

- development of a set of coherent and consistent archetypes,
- community review and approval of the archetypes as fit for use, using a collaborative peer-review process;
- publication and life cycle management of each archetype; and
- ongoing maintenance and governance of each archetype.

Each archetype is a computable specification for a single clinical concept – intended to be a maximal data set for a universal use case. In practice, the aspirational intent of a maximal data set is usually adjusted to a practical, but inclusive, data set that can be re-used across multiple clinical scenarios.

The program utilizes an online tool to support the activities of the program – the openEHR Clinical Knowledge Manager (CKM) [3]. This tool has three main purposes – it is a public library of archetypes; an open collaboration portal; and underpins the community's requirements for complex clinical knowledge governance.

The CKM tool allows open access to all clinical information models. Registration is required to actively participate within the CKM community – membership is free and open to any interested individual or group. A broad range of professions is represented including, but not limited to:

- Clinicians;
- Informaticians;
- Software engineers;
- Terminologists;
- Academics/students;
- Administrators; and
- Consumers.

As of December 18, 2016, the openEHR CKM has:

- 500 active archetypes in varying life cycle stages, comprising an estimated 6000 data points;

- 1628 registered users from 88 countries: and
- 24 languages represented.

Communities in Norway, Australia, United Kingdom, Slovenia, Canada and Brazil are actively collaborating and sharing archetypes to minimise 'reinventing the wheel'.

openEHR peer-review process

A small number of Clinical Knowledge Administrators are appointed to collectively take responsibility for the operations of the CKM instance. They appoint editors who are charged to develop and enhance the clinical content of each archetype from its initial draft through to a published state. The CKM tool supports this iteration by enabling the editors to run a series of review rounds to gather and collate reviewer feedback and manage the associated version and audit controls.

An archetype peer-review round is manually initiated by an Editor. They invite a subset of registered CKM reviewers - selected to ensure an appropriate cross section of professions, health domains and geographical location are represented. In addition, anyone who has a special interest in participating and requests an invitation to that review will also be included. Reviewing is optional and reviewers can opt out of any review invitation.

Review rounds are typically sent out for a period of two weeks. This review period can be adjusted by editors on a per review basis.

Reviewers can comment on any or all components of the archetype, guided through the various components by a 'wizard' process. They can also respond to some specific questions asked by the editor and targetting opinions on identified editorial issues. The only mandatory response is a final recommendation about readiness for publication:

- Accept – ready for publication;
- Minor Revision – trivial changes only (usually spelling/grammar), otherwise ready for publication without further community review;
- Major Revision – significant changes are needed, requiring further community review;
- Reject – not fit for publication or fundamentally flawed: and
- Abstain – no recommendation.

At the end of the review period the editors meet, usually via teleconference, to collectively respond to the feedback, update the archetype with agreed changes and decide on the next steps. If all recommendations are 'Accept' or 'Minor Revision' then consensus has been achieved and after the minor changes are applied, the archetype is ready to be published. If 'Major Revision' or 'Rejected' are recorded, then further review rounds are usually required until consensus is achieved.

All review comments plus the responses of the editors to each reviewer comment are captured as a record of provenance and viewable by all registered users, ensuring transparency of the decision process and accountability of the editors to the user community.

Approach

The following narrative outlines openEHR approach using a recent example of a complex cross-SDO collaboration between the openEHR and HL7 communities for

representation of Adverse Reaction Risk information models, also known as Allergy/Intolerance within the HL7 community. This narrative has been constructed retrospectively from audit trails and review round records captured within 3 CKM instances based the international openEHR community, Norway and Australia.

The very first iteration of the draft candidate for the Adverse Reaction archetype was authored in April 2006 by a single Australian clinical informatician, Dr Sam Heard. It was one of the first archetypes uploaded to the openEHR CKM [3] on 23 July 2008 [4]. In July 2009 this archetype commenced its' first collaborative peer-review in the openEHR CKM.

In November 2010, Australia's National eHealth Transition Authority (NEHTA, now known as the Australian Digital Health Agency [5]) forked the archetype and brought it into the Australian CKM [6,7] environment and ran a series of five archetype reviews during the period to June 2011. The resulting archetype content formed the basis for adverse reaction data points in CDA documents which are used to transmit health information from Australian primary care clinical systems into the PCEHR (now known as 'My Health Record [8]').

The Australian archetype formed the basis for a further iteration by Dr Heather Leslie which included feedback from international reviewers plus a variety of other resources including academic papers [9,10] and documents published and available at the time by NHS England [11,12], Microsoft's Clinical User Interface group [13,14], and the Royal Australian College of General Practitioners [15]. It was uploaded as a fork to the international CKM in January 2012 [16] and an international peer-review round was commenced.

In May and June 2014, further harmonisation by Dr Ian McNicoll merged feedback from the international review with content from HL7's FHIR resource and RMIM publications available at the time. In June 2014, due to the major structural changes it was uploaded as a new archetype [17] – the 'Adverse Reaction (FHIR/openEHR)' archetype with the intent of conducting a series of joint FHIR and openEHR reviews and generating both a FHIR resource AND an openEHR archetype with matching, clinically verified content at the end of the process.

In July 2014 the first joint openEHR/FHIR review was initiated. Four editors were appointed to facilitate the reviewer feedback – two from openEHR and two from HL7 and the resulting archetype was uploaded into the international CKM in October 2014 and a second review round initiated.

Concurrently, the Norwegian Nasjonal IKT team forked the archetype into the Norwegian CKM [18] in November 2014.

Resolution of the second openEHR/FHIR review round did not occur until June 2015, nearly 7 months later, due to delays caused by waiting for FHIR ballot results. This feedback was incorporated this feedback into the next archetype revision [19].

In June 2015 the third joint openEHR/FHIR review round commenced in the international CKM. Simultaneously, the archetype was updated in the Norwegian CKM, aligned the international archetype and translated to Norwegian [20]. Subsequently, in early August 2015, the first review round commenced in the Norwegian CKM. Feedback from this review was added to the international feedback so that parallel archetype development could evolve in English and Norwegian, each archetype revision now incorporating feedback from the international openEHR, Norwegian and HL7 communities.

The resulting archetype was uploaded to the international CKM in October 2015 [21] and the fourth openEHR/FHIR review round commenced. Soon after the second Norwegian review round commenced using the latest aligned and translated version of the archetype [22].

At the completion of the November 2015 review round and analysis of reviewer feedback, the editors agreed that a consensus about the archetype clinical content had been reached amongst the participating reviewers. In the openEHR archetype, all FHIR-specific components were removed and published as the 'Adverse reaction risk' archetype. The original archetype was rejected – this archetype persists in the international CKM as part of the provenance/audit trail for the published archetype but marked as not for current use.

During the review process, the FHIR team maintained an equivalent FHIR resource, adopting the changes agreed through the review process. This was the evolving artefact that was reviewed by the FHIR community. At the time of archetype publication, the content of the archetypes and the FHIR resource were aligned.

The Norwegian team updated their version of the archetype to align the content and the translation. They initiated a final review in Norwegian, commencing in late November 2015. At its conclusion the Norwegian archetype was also published within their local CKM and is now governed autonomously by the Nasjonal IKT team as per their national mandate. The agreed intent of the international and Norwegian teams is to continue to collaborate when change requests arise or new requirements are identified. These two openEHR archetypes remain semantically aligned as of December 18, 2016.

Results

It has been possible to collate review related data from each of the three CKM instances that have been used as part of the evolution of the Adverse Reaction Risk archetype through to publication.

A view of the review process is shown in Figure 1.

Discussion

This Adverse Reaction Risk archetype started its' journey as the brainchild of a single clinical informatician. After a journey of many twists and turns the final published archetype is the result of voluntary contributions from over 126 individuals, see Table 1, each contributing according to their professional background and expertise during 13 review rounds carried out in 3 CKM instances.

Table 1– Contributors statistics for each archetype

Archetype	Number of review rounds	Number of reviewers	Number of reviews
openEHR – initial	2	19	26
openEHR – openEHR/FHIR	4	38	69
NEHTA	5	37	66
Norway	2	32	42
Total	13	126	203

There has been no further formal joint collaboration between the openEHR and FHIR communities since the November 2015 publication. At that point in time, the great majority of clinical

content in both the resulting openEHR archetype and FHIR resource were aligned as a consequence of the joint review process.

Subsequently it appears that the FHIR resource has continued to evolve in isolation, effectively splitting from the jointly agreed artefact, apparently due to further requirements being identified in HL7 implementations [23]. This divergence is unfortunate, but unsurprising. It highlights that in order to achieve

cross-SDO standardisation of information models there will need to be a willingness to commit to an ongoing maintenance process as well.

Inclusion of the HL7 community was extremely evaluable in order to gather broader expert input and has no doubt improved the quality of the archetype. The final archetype was agreed in terms of the clinical content and then a pure openEHR archetype and a corresponding FHIR resource were developed, based on that common clinical content. This was a significant achievement, likely inevitable without a strong commitment from each party to maintain alignment.

From the openEHR point of view, the timing and frequency of the HL7/FHIR balloting process caused significant delays to the joint collaborative phase, resulting in an expected three to six months timeline for a complex information model blowing out to eighteen months. The frequent and short review cycles used by the CKM editors reflects a more agile and iterative approach targeting a single information model at a time.

The traditional SDO process is usually a closed activity in which value is placed on participation only by credentialed individuals, determined either by financial membership or nomination as an expert. By contrast, in the global Web 2.0 crowdsourced environment in which the openEHR communities of interest operate, the opposite conditions largely apply. The openEHR methodology places enormous weight on broad participation, accountability of those in roles of authority to every member of the community, and transparency at every level of governance:

1. **Participation is open and free** – participation is open to anyone who is willing to participate to the extent of their ability. It is not limited to individuals or organisations who have current paid memberships, who are nominated as 'experts', or who have been designated as 'credentialed' experts. This may be challenging to many but it supports input from the broadest professions, health domain expertise and geographical sources.
2. **Everyone can participate according to their expertise.** The user interface and review processes in the CKM tool has been developed specifically to ensure that non-technical experts, such as grassroots clinicians, can participate equally alongside the technology savvy. It removes the need for clinicians to acquire additional technical skills in order to participate. All feedback is encouraged, ranging from the smallest grammatical correction through to solutions for the most complex informatics or implementation conundrums.

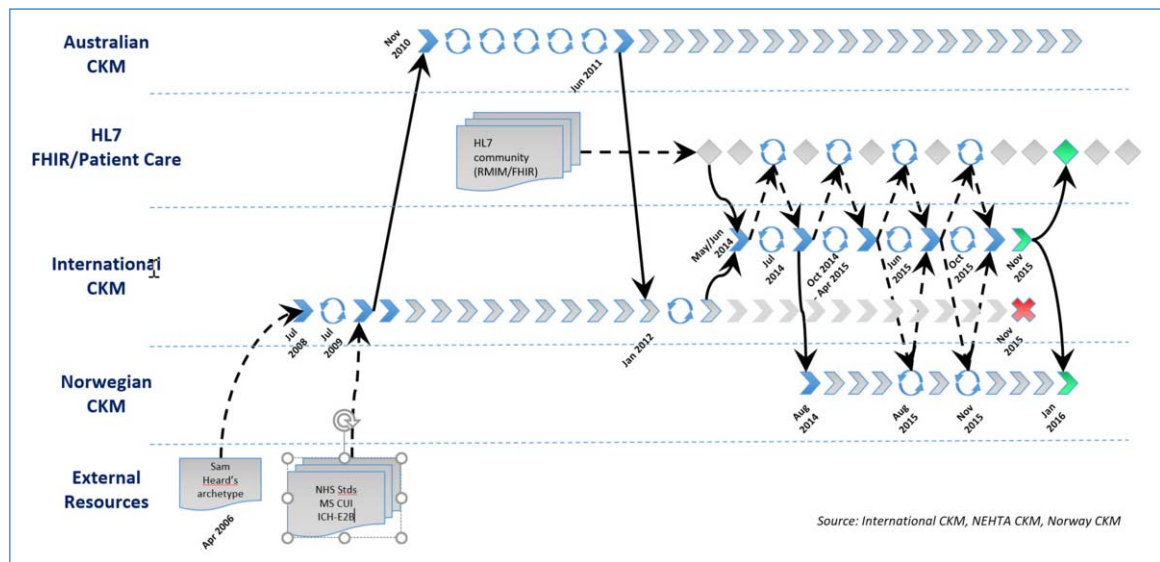


Figure 1 – Cross SDO Adverse Reaction Risk archetype collaboration process

3. Transparency. All of the activities and decision-making with CKM is transparent to registered users, including but not limited to:

- Archetype reviews, especially:
 - acknowledgement of all participants and their roles;
 - clear association between reviewer comments and resulting editorial decisions;
 - number of contributions; number of review rounds; and
 - composition of the reviewer community to ensure that an appropriate expert group has been involved.
- Threaded, unmoderated discussion threads;
- Change requests by registered users and editorial responses; and
- Archetype audit trail.

If a registered user is not happy with decisions there are a number of ways of raising this with editors or via public discussion boards.

4. Rapid and agile archetype publication. In the work that the openEHR Clinical Modelling Program have done to date, the typical archetype review process involves 4 review rounds to achieve broad agreement on the structure and data points. Sometimes further review rounds are required, usually focussed on refinement of archetype descriptions and metadata. With an average review round duration of two weeks, this means that an archetype requiring six review rounds could potentially be published in twelve weeks. Archetypes based on established and agreed clinical content such as evidence-based codes and scores can often be published in one or two review rounds – corresponding to between two and four weeks.

Assuming modest editorial resources are available, when multiple archetypes are being reviewed simultaneously it is possible to publish archetypes in efficient and effective timeframes.

By contrast, the traditional SDO ballot process would not be sustainable in the openEHR environment where the intent is to develop, review and publish all clinical archetypes required for all clinical data recording. There is a practical need for archetype review rounds to be:

- Managed as a sequence of short, frequent review rounds that result in progressively refined iterations of the archetype;
 - Initiated independently of other archetypes and for a variety of reasons, including initial publication, management of change requests and maintenance processes; and
 - Run when required - sometimes in parallel with other archetype reviews and at other times on an ad hoc basis to resolve a specific issue.
- 5. Shared archetypes amongst communities.** There are now a number of groups using the CKM tool as the basis of national or jurisdictional standardisation of data sets.

The traditional SDO process does not usually reveal the primary authors or contributors to their published standards, although they will possibly be known to SDO members. However the openEHR approach prioritizes transparency at every level of governance and for editors to be accountable to the CKM community:

- Free and open membership;
- Detailed audit trails to ensure accurate provenance and recording of editorial changes;
- Visibility of reviewer contributions and editorial responses
- Statistics about the review process, including:

- acknowledgement of all participants;
- number of contributions;
- number of review rounds; and
- background of all reviewers to ensure an appropriate reviewer community.

After publication of the adverse reaction archetype, collaboration between the openEHR CKM and Norwegian CKM teams has been active and ongoing. It has been successful largely because both groups are committed to working together and sharing the editorial work required to facilitate the reviewer feedback. openEHR and Norwegian editors meet regularly to collaboraton on solutions to modelling challenges, coordinate archetype reviews and update archetypes with feedback from both organisations. Reviews continue to be run in parallel in English and Norwegian on a range of archetypes - core content and specialised; simple and complex; crossing a broad range of clinical scenarios and professions. Archetype publication is based on the collective opinions of the communities that support both organisations. Both groups are willing for this to be extended to include other SDOs or national eHealth programs on request.

Conclusion

Clinical information modelling governance has been a new and largely untested challenge until recently – most of our collective experience in governance of health data standards has been at the complete message or document data set level. The Clinical Knowledge Manager tool was developed directly in response to identification of the need for efficient and responsive iterative refinement of the archetypes in response to identified requirements, especially during implementations – finding the sweet spot in the tensions between governance and evolution to ensure that the information models were safe and fit for use.

Clinical knowledge governance is a complex, evolving and poorly understood domain. The key to success of the openEHR approach, as described, is the result of humans choosing to collaborate to make a difference in the healthcare domain, using technology as the means to solve a shared problem. It is a socio-technical solution – a combination of the openEHR technical specification for an electronic health record architecture, a pioneering online knowledge governance tool, clinician engagement and a web 2.0 approach to harnessing the collective efforts of a community of volunteers.

Further analysis needs to be carried out to explore the impact of this approach as the number of published archetypes increases so that trends, patterns and conclusions can be identified.

Disclosure

The Clinical Knowledge Manager tool was developed by Dr Heather Leslie and Ocean Health Systems.

Address for correspondence

Dr Heather Leslie - heather.leslie@oceanhealthsystems.com

References

- [1] Clinical Models Program [Internet]. London: openEHR Foundation; [cited: 2016-12-23]. Available from: <http://www.openehr.org/programs/clinicalmodels/>
- [2] openEHR Foundation Home [Internet]. London: openEHR Foundation; [cited: 2016-12-23]. Available from: <http://www.openehr.org/>
- [3] openEHR Clinical Knowledge Manager [Internet]. London: openEHR Foundation; [cited: 2016-12-23]. Available from: <http://www.openehr.org/ckm/>
- [4] Adverse Reaction, Rejected Archetype [Internet]. openEHR Foundation, openEHR Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://www.openehr.org/ckm/#showArchetype_1013.1.197_1
- [5] Australian Digital Health Agency Home [Internet]. Sydney, Australia: Australian Digital Health Agency; [cited: 2016-12-23]. Available from: <http://www.digitalhealth.gov.au/>
- [6] Australian Digital Health Agency Clinical Knowledge Manager [Internet]. Sydney, Australia: Australian Digital Health Agency; [cited: 2016-12-23]. Available from: <http://dcm.nehta.org.au/ckm/>
- [7] Adverse Reaction, draft archetype, National eHealth Transition Authority [Internet]. NEHTA Clinical Knowledge Manager. Authored: 08 Nov 2010. Available at: http://dcm.nehta.org.au/ckm/OKM.html#showarchetype_1013.1.868_7 (accessed Jan 16, 2012).
- [8] My Health Record Home [Internet]. Sydney, Australia: Australian Digital Health Agency; [cited: 2016-12-23]. Available from: <https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/content/home>
- [9] Riedl MA, Casillas AM. Adverse drug reactions: types and treatment options. *Am Fam Physician*. 2003 Nov 1;68(9):1781-90. Review. PubMed PMID: 14620598.
- [10] Thien FC. Drug hypersensitivity. *Med J Aust*. 2006 Sep 18;185(6):333-8. Review. PubMed PMID: 16999678.
- [11] Horsfield P, Sibeko S. Representation in Electronic Patient Records of Allergic Reactions, Adverse Reactions, and Intolerance of Pharmaceutical Products [Internet]. London, UK: National Health Service; 2006 Sep 07 [cited: 2016-12-23]; Available at <http://webarchive.nationalarchives.gov.uk/+/http://www.isb.nhs.uk/documents/isb-1582/amd-24-2011/1582242011nfpitedb.pdf>.
- [12] Long R, Bentley S. SCG Guidance on the Representation of Allergies and Adverse Reaction Information Using NHS Message Templates [Internet]. London, UK: National Health Service; 2008 Apr 30 [cited 2011 Jun 21]; No longer available at <http://www.connectingforhealth.nhs.uk/systemsandservices/data/scg/scg0001.pdf>.
- [13] Microsoft. Design Guidance: Displaying Adverse Drug Reaction Risks [Internet]. 2009 January 28 [cited: 2016-12-23]; Available at <http://www.msui.net/DesignGuide/DisplayingAllergies.aspx>.
- [14] Microsoft. Design Guidance: Recording Adverse Drug Reaction Risks [Internet]. 2009 March 27 [cited: 2016-12-23]; Available at <http://www.msui.net/DesignGuide/Pdfs/Design%20Guidance%20--%20Recording%20Adverse%20Drug%20Reaction%20Risks.pdf>.
- [15] Royal Australian College of General Practitioners. Fact Sheet: Allergies & Adverse Reactions (Draft). 2010.
- [16] Adverse Reaction, Rejected Archetype [Internet]. openEHR Foundation, openEHR Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://www.openehr.org/ckm/#showArchetype_1013.1.197_5
- [17] Adverse reaction risk, Published Archetype [Internet]. openEHR Foundation, openEHR Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://www.openehr.org/ckm/#showArchetype_1013.1.1713_1
- [18] Nasjonal IKT Clinical Knowledge Manager [Internet]. Oslo, Norway: Nasjonal IKT HF; [cited: 2016-12-23]. Available from: <http://arketyper.no/ckm/>
- [19] Adverse reaction risk, Published Archetype [Internet]. openEHR Foundation, openEHR Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://www.openehr.org/ckm/#showArchetype_1013.1.1713_15
- [20] Risiko for overfølsomhetsreaksjon, Published archetype [Internet]. Nasjonal IKT, Nasjonal IKT Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://arketyper.no/ckm/#showArchetype_1078.36.579_2
- [21] Adverse reaction risk, Published Archetype [Internet]. openEHR Foundation, openEHR Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://www.openehr.org/ckm/#showArchetype_1013.1.1713_16
- [22] Risiko for overfølsomhetsreaksjon, Published archetype [Internet]. Nasjonal IKT, Nasjonal IKT Clinical Knowledge Manager [cited: 2016-12-23]. Available from: http://arketyper.no/ckm/#showArchetype_1078.36.579_5
- [23] Graham Grieve. FHIR and openEHR [Internet]. HL7 International; 2016 Nov [cited: 2016-12-23]. Available from: <https://vimeo.com/user12740828/hl7-fhir-developer-days-2016-amsterdam/video/191939835>

A Semi-Automatic Framework to Identify Abnormal States in EHR Narratives

Xiaojun Ma^a, Takeshi Imai^a, Emiko Shinohara^b, Ryota Sakurai^b, Kouji Kozaki^c, Kazuhiko Ohe^{a,b}

^a Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

^b The University of Tokyo Hospital, Tokyo, Japan

^c The Institute of Scientific and Industrial Research, Osaka University, Osaka, Japan

Abstract

Disease ontology, defined as a causal chain of abnormal states, is believed to be a valuable knowledge base in medical information systems. Automatic mapping between electronic health records (EHR) and disease ontology is indispensable for applying disease ontology in real clinical settings. Based on an analysis of ontologies of 148 chronic diseases, approximately 41% of abnormal states require information extraction from clinical narratives. This paper presents a semi-automatic framework to identify abnormal states in clinical narratives. This framework aims to effectively build mapping modules between EHR and disease ontology. We show that the proposed method is effective in data mapping for 18%-33% of the abnormal states in the ontologies of chronic diseases. Moreover, we analyze the abnormal states for which our method is invalid in extracting information from clinical narratives.

Keywords:

Ontology; Machine learning; Natural language processing

Introduction

In order to understand diseases, one must adequately capture abnormal states in diseases. With support from Japan's Ministry of Health, Labour and Welfare, we have been involved in developing a disease ontology [1,2], wherein a disease is captured as a causal chain of abnormal states. So far, medical experts have described the causal chains of approximately 6,000 diseases across 13 medical departments. The disease ontology provides domain-specific knowledge, answering questions such as "what abnormal states cause a disease?" and "how might a disease progress, and what symptoms may appear?". We believe that disease ontology is a valuable knowledge base in medical information systems such as deep phenotyping in support of precision medicine. For the practical use of disease ontology in clinical applications, automated mapping between electronic health records (EHR) and disease ontology is indispensable. An investigation has been conducted to specify clinical data sources where information can be extracted to identify abnormal states. The investigation is based on an analysis of 4,718 (643 distinct) abnormal states of 148 chronic diseases. We categorize data sources in EHR as clinical notes, exam reports, laboratory data, treatment orderings, and demographics, which each require different mapping techniques. From the investigation results, approximately 41% of the abnormal states are supposed to be identified in unstructured clinical notes or exam reports. Therefore, our study first explores techniques of mapping clinical narratives to abnormal states, which involves machine learning and natural language processing (NLP).

In previous studies, Informatics for Integrating Biology and the Bedside (i2b2) has organized NLP challenges including identifying patient smoking [3] and obesity [4] status from English-language discharge summaries. A common method exploited by the participants can be summarized in two steps: (1) retrieve relevant sentences or passages via keyword or rule-based search, (2) build sentence-level or passage-level classifiers to determine status, which we refer to here as polarity classification. In this paper, we present a general framework, extended from the conventional method, for extracting information from clinical narratives to identify abnormal states in disease ontology. The difference between our study and the i2b2's challenges lays in the following aspects: (1) all available clinical notes in EHR are used in our study including discharge summaries, progress notes, nursery notes, which incorporate a wide variety of narratives; (2) the clinical narratives in our study are written in Japanese, which requires Japanese language processing such as morphological analysis. The improvements we have made to the method include the following: (1) adopting a semi-automatic keyword expansion for candidate sentence retrieval, using both knowledge-driven and data-driven approaches; (2) applying Synthetic Minority Over-sampling Technique (SMOTE) [5] to solve class imbalance problem in polarity classification; and (3) introducing a mechanism of using existing labeled data to automatically label unlabeled data, which reduces the labeling costs of machine learning for polarity classification.

Our study focuses on chronic diseases across the life span including diabetes, hypertension, dyslipidemia, chronic kidney disease, and ischemic heart disease. The objective of our study is to propose a general framework to efficiently build mapping modules between clinical narratives and abnormal states in ontologies of chronic diseases. In this paper, we discuss the technical requirements to identify abnormal states in clinical narratives, and we validate the applicability of the proposed framework using EHR data from the University of Tokyo Hospital. Experiments demonstrate that the proposed method is effective in EHR data mapping for 18%-33% of the abnormal states in the ontologies of chronic diseases.

The rest of the paper is organized as follows. The following section introduces the results of the investigation into abnormal states. The Methods section describes the proposed framework. The Experiment Section describes experiments designed to evaluate the performance. In the Discussion section, we discuss the proposed method and those abnormal states whose mapping modules cannot be realized by our framework. The final section presents our conclusions.

tokens beforehand, among which nouns are used to train word2vec models. The JMA is based on a customized medical dictionary.

Polarity Classification

For each retrieved sentence, affirmation or negation (polarity) of an abnormal state is determined. It is a classification problem, where three methods are applied to build classifiers.

Keyword Look-up (KL)

With respect to some keywords, their presence can directly indicate polarity. For example, the existence of “bi=” (“bi” is an abbreviation of “brinkman index”) implies that the patient is a smoker.

Rule-based Algorithm (RA)

Regular expressions are created to determine the polarity of sentences retrieved by some keywords, especially English keywords such as “smoker”, “smoking”, and “obesity”. In the case of abnormal states of disease names such as “diabetes”, and “dyslipidemia”, it is rare to see their negations and the negative expressions are quite similar across diseases. Therefore, limited negative expressions can be collected for polarity classification and shared by multiple disease names.

Machine Learning (ML)

Sentences related to symptoms and those related to lifestyle issues retrieved by Japanese keywords show great variety in the contexts in which they appear. In these cases, machine learning is applied.

We use linear support vector machine (SVM) as a base algorithm and bag-of-words (BoW) as features. To construct a BoW feature for each sentence, the JMA is conducted to tokenize each sentence into words. The window size of words used to construct features is adjusted for the best performance.

For an abnormal state, affirmative sentences are often much more common than negative ones. For those data with severe class imbalance problem, SMOTE is applied.

Input

$\{s_i\} \in S$: set of labeled data in source domain

$\{t_j\} \in T$: set of unlabeled data in target domain

Output

$T^l \in T$: set of labeled data in target domain

Procedure

1. Train a KNN model on S .

2. Use the model to annotate each $\{t_j\} \in T$ to get h_j : $t_j \rightarrow$ label by majority voting of the most k closest s_i with probability. For $j = 1, \dots, |T|$ do:

If $\text{prob}(h_j) = 1$ then

Set label of t_j

with h_j

$T^l = T^l \cup t_c$

Figure 3 – Mechanism for RLC

Labeling data is expensive. We try to reuse existing labeled data of an abnormal state (source domain) to reduce the workload of labeling unlabeled data for a new abnormal state (target domain). We establish a mechanism for reducing manual labeling, denoted as RLC, as described in Figure 3. A k-nearest neighbors (KNN) model trained on labeled source-domain data is applied to annotate unlabeled target-domain data. The labels outputted with high probability are assigned to the data in the target domain automatically, while the rest of the unlabeled data remain for manual labeling

Experiments

Data Description

The data for this study is based on EHR data from the University of Tokyo Hospital, involving a great variety of types of clinical notes, including discharge summaries, progress notes, nursery notes. Exam reports are usually transcribed in clinical notes, so the data we used includes texts of both clinical notes and exam reports.

Dataset 1 (DS1)

We collected all clinical notes generated in 2015 at the University of Tokyo Hospital, which is denoted as DS1. DS1 is composed of approximately 170,000 clinical notes, which we used to evaluate our method.

Dataset 2 (DS2)

In order to train word2vec model for keyword expansion, we collected a dataset denoted as DS2. DS2 consists of approximately 60,000 clinical notes generated from 2010 to 2015 at the University of Tokyo Hospital, which contain at least one of the following disease names: chronic kidney disease, is-chemic heart disease, and arteriosclerosis.

Experimental Settings and Evaluation Methods

Two things we concern are: first, “how precise is identification of an abnormal state”, and second, “how many patients with a given abnormal state are recalled”.

Evaluation of polarity classification answers the first question. We constructed a KL or RA by viewing approximately 20 candidate sentences. For each KL or RA classifier, if the total number of candidate sentences in DS1 is much higher than that used for classifier construction, we sampled approximately 100 sentences to create test data for evaluation. To construct a training set for a ML, we randomly sampled 250 sentences, if available, from candidate sentences retrieved from DS1. For those abnormalities for which there were not enough candidate sentences, all of the available sentences were used. The training sets were manually labeled and five-fold cross validation was conducted to evaluate the performance of ML classifiers.

A direct answer to the second question is achieved by computing patients’ recall for every abnormal state. This requires annotation by medical experts, which is extremely expensive in terms of both time and financial cost. In our proposed method, recall depends on the effectiveness of keyword expansion. Therefore, at this stage, we present increased ratios of retrieved clinical notes by keyword expansion instead.

Experimental Results

We applied our proposed framework to 10 abnormal states, which are summarized in Table 3.

Effectiveness of Keyword Expansion

The keyword expansion grew the keyword lists for candidate sentence retrieval. Table 3 shows the increased percentages of retrieved notes from DS1 compared to those retrieved before keyword expansion (see column “Eva on KE”).

Performance of Polarity Classification

Table 3 gives the methods of polarity classification for each abnormal state and their performance measured in macro-averaged precision (Avg. P), recall (Avg. R), and F1 score (Avg. F1), as well as classification accuracy (CA).

For ML, we investigated the effect of dataset size on macro-averaged F1 scores of five-fold cross validation (see Figure 4). We found that for most abnormal states with dataset size above 100-250, F1 increases more slowly and standard deviation of

F1 converges. The observation is basically consistent with a previous study [10] of the i2b2 Smoking Challenge, which states that accuracy increases slowly when the training set size is over 200 using hold-out validation. Therefore, in this study, we used 250 labeled sentences for each ML (except for the abnormal state of pulmonary congestion which has only 142 candidate sentences available in DS1). The performance shown in Table 3 for ML is based on fully manually annotated data without introducing RLC.

Table 3 – Effectiveness of keyword expansion and performance of polarity classification

Abnormal States	Eva on KE	Polarity Classification				
		Methods	Avg. P	Avg. R	Avg. F1	CA
diabetes	52.0%	RA	-	-	-	0.98
dyslipidemia	68.6%	RA	-	-	-	0.99
obesity	37.9%	KL, RA	1.00	1.00	1.00	1.00
smoking	1.1%	KL, RA, ML	0.94	0.95	0.94	0.95
anura	0.0%	KL	-	-	-	-
chest pain	52.7%	ML	0.93	0.95	0.94	0.95
fever	36.7%	RA, ML	0.93	0.97	0.95	0.96
overhydration	41.4%	ML	0.89	0.91	0.90	0.91
pulmonary congestion	29.3%	ML	0.90	0.92	0.91	0.92
tachycardia	186.2%	ML	0.95	0.96	0.96	0.96

* Negation of a disease name is extremely rare, so only CA is given.

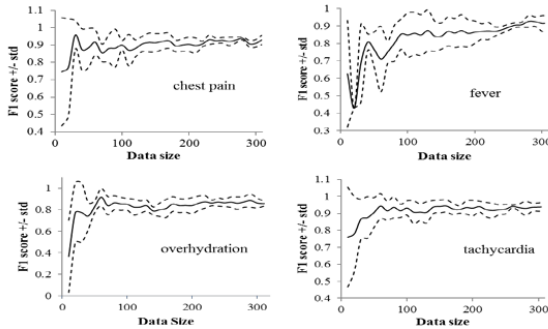


Figure 4 – Effect of dataset size on F1 scores

Availability of RLC

In the experiments on polarity classification using the fully manually annotated data, ML models for chest pain, fever, overhydration, pulmonary congestion and tachycardia showed the best performance with the same word count window in feature construction, which uses eight tokens after the keywords. All these abnormal states are symptoms. We assume that their candidate sentences may have similar contexts. A new experiment was designed to verify RLC based on these five abnormal states of symptoms.

In order to measure similarity between feature spaces of retrieved sentences containing the five symptoms, we computed the cosine similarity between word count vectors of the 250 annotated sentences for every two symptoms, which is shown in Table 4. Table 5 describes the performance (F1 scores) of the ML models trained on data of one symptom to predict polarity for the other four symptoms. The Pearson correlation coefficient between Table 4 and Table 5 is 0.422, indicating moderate positive correlation.

We applied RLC to the five abnormal states of symptoms, using labeled data of one symptom (source domain) to reduce the labeling costs of the other four (target domain). The rates of reduced labeling costs, that is, the proportions of automatically labeled data in the target domains, are shown in Table 6, ranging from 1.6% to 69.7%, averagely 32.4%. Table 7 gives the macro-averaged F1 scores of five-fold cross validation of

polarity classification using the RLC outputs, which is similar to the performance using the fully manually annotated data as shown in Table 3.

Table 4 – Cosine similarity between word count vectors for every two symptoms

	chest pain	fever	over-hy- dration	pulmonary congestion	tachy- cardia
chest pain	-	0.89	0.76	0.82	0.91
fever	-	-	0.79	0.84	0.85
overhydration	-	-	-	0.79	0.83
pulmonary congestion	-	-	-	-	0.82
tachycardia	-	-	-	-	-

Table 5 – Performance of polarity classifiers trained on data of one symptom to predict others

Train \ Test	chest pain	fever	over-hy- dration	pulmonary congestion	tachy- cardia
chest pain	-	0.87	0.73	0.88	0.9
fever	0.92	-	0.77	0.86	0.93
overhydration	0.9	0.88	-	0.9	0.89
pulmonary congestion	0.92	0.87	0.74	-	0.91
tachycardia	0.9	0.88	0.73	0.89	-

Table 6 – Percentage of reduced labeling costs

Source \ Target	chest pain	fever	over-hy- dration	pulmonary congestion	tachy- cardia
chest pain	-	33.6	16	19.7	42
fever	43.6	-	40	49.3	50
overhydration	13.2	14	-	12	21.2
pulmonary congestion	11.6	10	1.6	-	14
tachycardia	61.2	63.6	60.8	69.7	-

Table 7 – Performance of polarity classification using RLC

Source \ Target	chest pain	fever	over-hy- dration	pulmonary congestion	tachy- cardia
chest pain	-	0.96	0.88	0.90	0.96
fever	0.94	-	0.86	0.92	0.97
overhydration	0.93	0.96	-	0.91	0.96
pulmonary congestion	0.93	0.96	0.86	-	0.96
tachycardia	0.95	0.96	0.89	0.95	-

Discussion

Applicability of the Proposed Framework

Our proposed framework is useful in mapping clinical narratives to abnormal states that are described by word-level expressions and have concrete clinical meanings. Keyword expansion and polarity classification are two key processes in our method.

Table 3 shows that the keyword expansion increases the number of retrieved clinical notes by 51% on average, which demonstrates the effectiveness of keyword expansion in improving recall. We also found that word2vec is able to find candidate keywords that the dictionary-based method cannot obtain, such as orthographical variants, obsolete terms but still in use, and informal expressions used locally in each hospital. However, keyword candidates outputted by word2vec show low precision, which requires manual selection. In future work, refining word representation learned from corpora by using semantic lexicons [11] may be a promising approach to outputting keyword candidates more efficiently.

As for the polarity classification, the proposed framework solves two problems in building ML classifiers: class imbalance and labeling costs. We applied SMOTE to solve the former problem, improving the macro-averaged F1 scores of cross validation by 5% to 10%. The F1 scores with SMOTE applied

are shown in Table 3. For the latter problem, we introduced RLC. Figure 4 shows that the size of the training data set necessary to build a high-performance ML classifier is above 100-200. Labeling training data for every abnormal state is time-consuming. RLC reuses existing labeled data to automatically label those unlabeled data that are similar to the existing labeled ones, thus reducing labeling costs. From Table 4 and Table 5, when the training and the test data come from different domains, moderate positive correlation exists between the cosine similarity of their feature spaces and the performance of the models. It is reasonable to consider using data in source domain to annotate data in target domain if they have similar feature spaces. Based on the experiments on the five abnormal states of symptoms, Table 6 and Table 7 demonstrate that RLC can reduce the labeling costs by 32.4% on average, achieving similar performance as using the fully manually annotated data. The rate of reduced labeling costs possibly depends on the dataset size of the source domain, and the similarity between feature spaces of the source and target domain. RLC is a simple and effective way of reducing labeling costs in our proposed framework.

Limitations of the Proposed Framework

Our proposed framework is useful in mapping clinical narratives. The proposed framework is invalid in identifying some abnormal states whose information is supposed to be contained in clinical narratives. We summarize these kinds of abnormal states as follows.

Some abnormal states are expressed at sentence level in EHR like “feel pain in chest”, and “the spleen is enlarged”. These kinds of expressions are missed in candidate sentence retrieval which is based on keyword search in our proposed method.

Abnormal states like “stress” and “systematic inflammation” have abstract clinical meanings, which causes difficulty in collecting keywords for retrieving candidate sentences.

Some abnormal states such as “exertion” usually appear with other states such as “chest pain upon exertion”, and “difficult respiration upon exertion”. Thus, retrieved sentences do not focus on “exertion” but respond to multiple abnormal states. Therefore, mapping module should be established between EHR and a cluster of abnormal states in disease ontology.

Extraction of numerical information such as body weight, body mass index (BMI), and blood pressure is necessary for identifying some abnormal states like “obesity” and “hypertension”. Numerical extraction from narratives requires more sophisticated NLP techniques.

Future Directions

Our future work will include the following: (1) building modules of mapping other data sources such as laboratory data and treatment orders to disease ontology; (2) developing sophisticated NLP techniques to identify the abnormal states in narratives that cannot be achieved by the proposed method; (3) exploiting structures in disease ontology to infer those abnormal states that have no trace of evidence in EHR.

Conclusions

In this paper, we have proposed a general framework to identify abnormal states in EHR narratives, which is applicable in data mapping for 18%-33% of the abnormal states in the ontologies of chronic diseases. The proposed method exhibits improvement over the conventional method by applying keyword expansion based on UMLS and word2vec, and by introducing SMOTE and RLC in ML for polarity classification.

Acknowledgements

This research is partially supported by the ICT infrastructure establishment for clinical and medical research from Japan Agency for Medical Research and development, AMED, and Health and Labour Sciences Research Grants from the Ministry of Health, Labour and Welfare, Japan.

References

- [1] K. Kozaki, R. Mizoguchi, T. Imai, and K. Ohe, Identity tracking of a disease as a causal chain, *Proc. ICBO* (2012), 131-136.
- [2] K. Kozaki, R. Mizoguchi, T. Imai, and K. Ohe, A consideration on identity of diseases. *Proc. InterOntology* (2012), 75-80.
- [3] O. Uzuner, I. Goldstein, Y. Luo, and I. Kohane, Identifying patient smoking status from medical discharge records, *JAMIA* **15** (2008), 14-24.
- [4] O. Uzuner, Recognizing obesity and comorbidities in sparse data, *JAMIA* **16** (2009), 561-570.
- [5] N.V. Chawala, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, *JAIR* **16** (2002), 321-357.
- [6] Unified medical language system (UMLS). Available at: <https://www.nlm.nih.gov/research/umls/>. [accessed Dec. 2016]
- [7] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, Distributed representations of words and phrases and their compositionality, *NIPS* (2013), 3111-3119.
- [8] Merck Manual Professional Edition in Japanese. Available at: <http://merckmanual.jp/mmpej/index.html>. [accessed Dec. 2016]
- [9] Merck Manual Home Edition in Japanese. Available at: <http://merckmanuals.jp/home/index.html>. [accessed Dec. 2016]
- [10] C. Clark, K. Good, L. Jezierny, M. Macpherson, B. Wilson, and U. Chajewska, Identifying smokers with a medical extraction system, *JAMIA* **15** (2007), 36-39.
- [11] M. Faruqi, J. Dodge, S.K. Jauhar, C. Dyer, E. Hovy, and N.A. Smith, Retrofitting word vectors to semantic lexicons, *Proc. NAACL-HLT* (2015), 1606-1615.

Address for correspondence

Xiaojun Ma, Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655 Japan.

Email: xiaojun_ma@m.u-tokyo.ac.jp

Application of Ontology Technology in Health Statistic Data Analysis

Guo Minjiang, Hu Hongpu, Lei Xingyun

Institute of Medical Information/Center for Health Policy and Management, Chinese Academy of Medical Sciences, Beijing, China

Abstract

Research Purpose: establish health management ontology for analysis of health statistic data. *Proposed Methods:* this paper established health management ontology based on the analysis of the concepts in China Health Statistics Yearbook, and used protégé to define the syntactic and semantic structure of health statistical data. *Results:* six classes of top-level ontology concepts and their subclasses had been extracted and the object properties and data properties were defined to establish the construction of these classes. By ontology instantiation, we can integrate multi-source heterogeneous data and enable administrators to have an overall understanding and analysis of the health statistic data. *Conclusion:* ontology technology provides a comprehensive and unified information integration structure of the health management domain and lays a foundation for the efficient analysis of multi-source and heterogeneous health system management data and enhancement of the management efficiency.

Keywords:

Disease Management; Semantics; Models, Statistical

Introduction

As the Chinese health system reform is deepened constantly, to analyze the statistical data of health care system efficiently becomes more and more important for decision-makers to formulate policies, assess the progress and predict trends.. However, the increasing number of terms used in the health statistic data leads to challenges for information staff. It is difficult for them to achieve an adequate understanding of the concepts denoted by the terms and to grasp the concepts' interrelations, which hinders data analysis [1].

Ontology is the part of the information system that explicitly commits it to a certain conceptualization of the world [2, 3]. In the health sciences, ontological techniques are applied towards a variety of goals among knowledge modeling, organization, integration, and exploitation for studies about genomics, anatomy and so on [4]. It can present the scenario from the perspective of the service user, from that of the overall organization, or from that of a sample to be analyzed. Many ontology work has been done in this field. The Unified Medical Language System (UMLS) which was developed by the National Library of Medicine (NLM), is an information retrieval ontology. UMLS is a set of files and software that bring together many health and biomedical vocabularies and standards to enable interoperability between computer systems[5]. The Open Biomedical Ontology (OBO) is a biomedical data ontology, which is a collection of ontologies, such as gene ontology, cell type, plant ontology and sequence

ontology. OBO makes the data generated in biomedical research to form a single, consistent, cumulative expansion and easy to operate on the algorithm as a whole [6]. GALEN is a clinical information ontology, which produces a computer-based multilingual coding system for medicine. The ontology is divided into 2,738 modules, which are organized according to the tertiary tree structure, and the relationship is grouped together [7]. The existing ontology research work provides an effective tool for integrating large amounts of medical data. Although health management data is different from the medical data, these types of research studies still provide insights for organizing health management data.

In consideration of the above problems, based on a systematic analysis of core concepts in health statistic data, we have constructed a system that focuses on the use of data from China Health Statistics Yearbook combined with health management ontology. This article describes the conceptual model of the ontology and defines the syntactic and semantic structure of health statistical data, as well as the terminologies generally accepted by health administrators. Through the construction and application of the health management ontology, it aims to integrate and share various health management data resources in the operation of the health system.

Methods

Materials

Health information resources are featured with interdisciplinary, large, and low-standardized vocabulary. Currently, the ontology for management is usually constructed according to different features of the respective disciplinary domain and the specific project, and there is not a systematic and engineered ontology construction method. A lot of management ontology construction methods are summarized from a single project or case [8]. On one hand, the size and scope of the ontological concept cannot fully meet requirements of health administrators for operating management of the health system. On the other hand, the ontology constructed can hardly complete an overall analysis of the whole operating process of the health system. Health statistical yearbooks serve as an important basis for providing a comprehensive understanding of the health system status, determining tasks and objectives of the health work, and implementing decisions of the health management and reform. Its indexes best embody the core concept, object and relevance of health system management and are suitable for constructing a comprehensive and systematic ontology of the domain of health management.

In order to understand the themes and structures in health management, we established a data model based on the star model, and used a text similarity calculation method to identify the dimensions and correlations of indicators. Data structure and relationship of health statistics yearbooks were expressed in the model. Therefore, taking the index and dimension names in the yearbooks as the text source, this research extracted the core objects and relationships in the health statistic data to integrate and analyze the operating information resources of the health management.

Methods

There are many methods of ontology construction, such as the skeletal methodology [9], TOVE methodology [10], METHONTOLOGY methodology [11], 7-step methodology [12], etc. By referring to these methods, this article makes some proper adaptations according to the features of the application requirements of this domain. Its general strategy is to use the top-down (first, design the top-level concepts, then provide detailed, step-by-step information) and bottom-up (first, collect domain concepts from domain documents as many as possible, then generalize them step by step) method. The specific construction flow is indicated in Figure 1, mainly including the demand analysis, resource reusing, building the ontology framework, adding instances, formalizing the ontology and ontology evaluation.

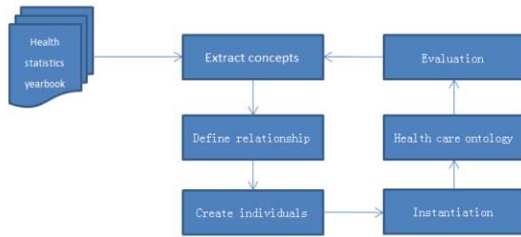


Figure 1—Health Management Ontology Construction Flow

TABLE 1-DIMENSION TABLE EXAMPLE

Dimension Code	Category Code	Dimension Name	Dimension Category Name
001-01	001	Hospitals	Health institutions
001-02	001	Primary health institutions	Health institutions
001-03	001	Professional public health institutions	Health institutions
001-04	001	Other health institutions	Health institutions

Concept Extraction

We took the China Health Statistical Yearbooks in 2004-2010 as an example, which are divided into different topics as required by health management. In reference to the UMLS and OBO ontology system, we followed two principles to extract the concepts of the top-level ontology. One is the peer principle, which means that concepts of the top-level ontology have the consistent scope of expression. The other is the relevance principle, which means that different concepts are related to each other objectively. On these bases, we defined the topics with a wide scope of coverage in the “yearbooks” as the top-level ontology concepts; and we combined the topics with narrow scopes of coverage into a new topic as the top-level ontology concept; topics reflecting the health management efficiency indexes, such as the level of people’s

Tools

At present, optional mature tools for the ontology development both home and abroad include Protégé, OntoEdit, WebODE, etc. In consideration of the usability and openness, this research uses Protégé 5.0 for development and chooses OWL as the language to describe the administrative ontology of the health system.

Results

By analyzing the concepts, objects and relationships involved in the health statistical yearbooks, this article defines the concepts, terminologies and their properties presented in the health system.

Data cleansing

First, we built a star data model for the statistical data. Every table in the yearbook is split per its properties and by indicator, time, geographic region and dimension, to form a star schema centered on “Yearbook data table” by dimensional tables.

Second, it was necessary to determine the correlation among indicators to extract the concept and relationship. Since the indicator name and dimension name are text information, we applied a text similarity calculation method to automatically identify the “dimensions” and the “upper level indicators”. A positive maximal matching principle of the semantic analysis method was applied to calculate the similarity. Then the created “indicator table” and “dimension table” were utilized to determine the dimension of an indicator via a custom macro program to find out its corresponding upper level indicators.

After that, a star schema structure was finally created, in which, the concept and the relationship is clearly defined (table 1).

health and causes of residents’ disease, injury and death, were not defined as the top-level ontology concepts. Upon the aforesaid analysis, we extracted six classes of top-level ontology concepts in the health management domain, which are the health institutions, health personnel, health facilities, health funds, health services, and health insurance.

Terminology Definition

To define the conceptual vocabulary properly is a key process in ontology construction. Only a good conceptual hierarchical structure can describe the knowledge structure of the health system clearly and accurately. Based on the top-level ontology concepts defined above, we classify their underlying relationships and terminologies hierarchically. We used some class of Ontology of Medically Related Social Entities from OBO, but in order to incorporate it with the Chinese health management system, some class names have been modified.

Healthcare provider organizations refer to the organizations providing the society with services of medical health care, disease control and health inspection or are engaged in medical scientific research and on-the-job medical training, for example, hospitals, basic health institutions, and special public health institutions. Healthcare providers refer to people working in healthcare provider organizations for medical treatment, prevention and healthcare, medical scientific research, and in-service education, for example, doctors, nurses, pharmacists, medical technicians, administrators and logistics workers. Health facilities refer to the facilities used in health institutions to provide the society with various health services, for example, houses, beds, and medical equipment.. Health funds refer to incomes and expenses in the operating process of the health system, for example, medical incomes and pharmaceutical incomes. Health services refer to medical health services provided by various health institutions for the society, for example, medical services, public health services, and women and children healthcare. Medical security systems refer to systems providing security for the medical health services for all populations, for example, the new rural cooperative medical care system, the basic medical insurance system for employees in urban areas, and the basic medical insurance system for urban residents.

Figure 2– Health Management Ontology Class

The ontology defines the property to define the class relationship, which is implemented by setting the domain and range of the properties. There are mainly two properties. One is the object property, used to relate an individual to another; the other is the data property, used to relate an individual to the data type.

In order to reflect the structures and rules existing in different concepts of the health system, we also needed to describe the class relationships after concept extraction. There are mainly six relationships:

- “inheres in” is used to describe the relationship between “health care provider” and the “health care provider organization”.
- “is bearer of” is used to describe the relationship that a “health care provider organization” bears a “health care provider”.
- “Earn” is used to describe the relationship that a “health care provider organization” or a “person” earns the “health fund”.
- “Expense” is used to describe the relationship that a “health care provider organization” or a “person” or “health service” expends the “health fund”.
- “Provide” is used to describe the relationship that a “health care provider organization” or “health care provider” provides a “health service”.
- “Insure” is used to describe the relationship that a “health security system” insures a “health care provider organization” or “health service”.

Data Properties

The object properties and data properties of the health management ontology created in this article are indicated in Figure 3 and Figure 4.

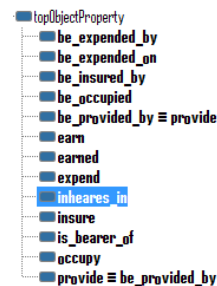


Figure 3—Health Management Ontology Object Properties

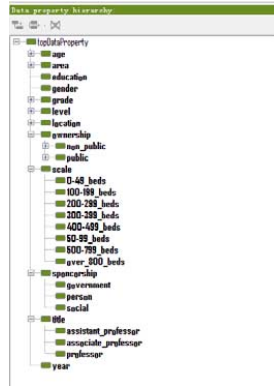


Figure 4– Health Management Ontology Data Properties

Ontology Instantiation

After defining the class and property structure of the health system, we can instantiate the ontology, which can be implemented by a manual entry, machine import and artificial semantic annotation. In this article, we input three instances via the manual entry window of protégé: a general hospital “Peking Union Medical College Hospital”, a doctor “Zhang San”, and equipment “MRI”. The relationships of the three instances are indicated automatically in ontograp as shown in Figure 5. We can see the line combined “Peking Union Medical College Hospital” and “MRI” show the relationship: Zhang San – work in - Peking Union Medical College Hospital.

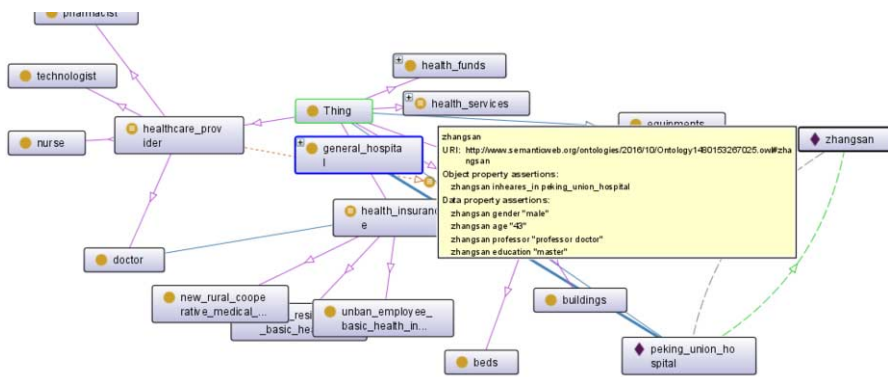


Figure 5– Health Management Ontology Instantiation

Application

By ontology instantiation, we can integrate multi-source heterogeneous data and enable administrators to have an overall understanding and analysis of the health statistic data. Figure 6 shows the operation interface of data analysis upon the construction of the integrated data of the health

management domain ontology and establishment of the “health data analysis platform”, which can also be established by relational database but requires a lot of data processing work. By clicking any index on the figure, we can be linked to its related instance and recompose the figure for the convenience of classification and aggregation of the health system data.

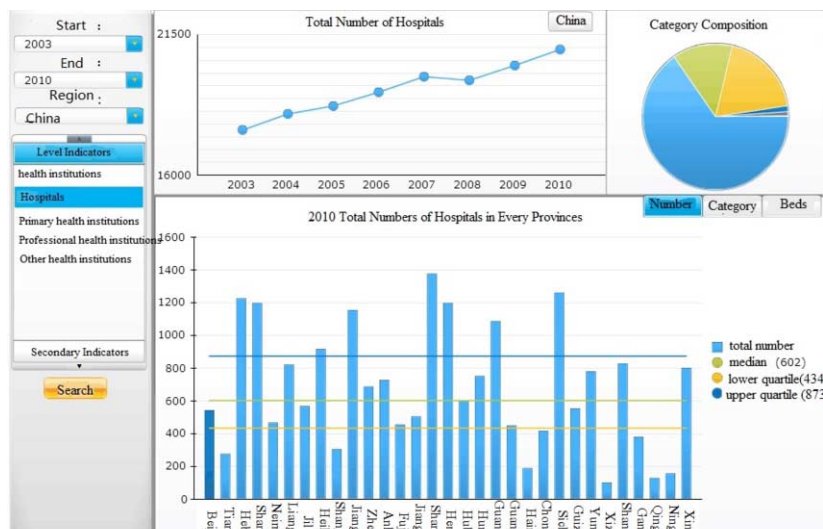


Figure 6– Application of Health Management Ontology in Health Data Analysis Platform

Discussion

Using the index and dimension names in the health statistical yearbooks as the text sources, this article constructs the domain ontology of the health system, defines the concepts of top-level ontology, core terminologies object properties and data properties, applies the ontology to the “health data analysis platform” to facilitate the aggregation and classification analysis of the platform data, and lays a foundation for the final implementation of the domain’s special knowledge bases based on the ontology. However, as the data of the health system have a wide scope and limited degree of openness, currently, the instance data still need to be expanded. Subsequent work includes further improvement of the ontology by ontology learning, ontology evolution by dynamic updating and automatic expansion of special knowledge bases.

With the continuous openness of health data, there will be huge amount of health data for management and policy research, including new organization types, provider types, insurance types and so on, which will drive the ontology to evolve. Use of the health management ontology can also be employed in other health management information systems by rapidly integrating different sources and heterogeneous data and helping supervisors to quickly grasp the development trend, and make decisions.

Conclusion

As the age of precision medicine is approaching, data become more and more important for the development and prediction of the health industry. It becomes more necessary to have a deep understanding of historical data by intelligent analysis so as to know the industry status and predict development trends accurately. To construct an appropriate ontology structure is the basis for efficient and comprehensive intelligent analysis. Taking China Health Statistical Yearbook as an example, by extracting various concepts and relationships, this article forms a comprehensive and unified information integration structure of health management domain and lays a foundation for efficient analysis of multi-source and heterogeneous health system management data and enhancement of management efficiency. In the future, combining with the industrial requirements of health system management and the analysis methods of big data, we will dig deeper into the data and fully exert the supporting role of informatization in the development of the health cause.

Acknowledgements

This work is supported by special fund for basic scientific research business of central public research institutes (2016RC330017).

References

- [1] Peter H, John H, Xinzi W et al. Ontology-Based Tools to Expedite Predictive Model Construction. *AMIA Jt Summits Transl Sci Proc*. 2014. Published online 2014 Apr 7.
- [2] Gruber TR. Toward principles for the design of ontologies used for knowledge sharing [J]. *Int J Hum Comput Stud*. 43 (1995), 907–28.
- [3] Guarino N. Formal ontology in information systems In: Guarino N, editor. *Proceedings of the Second International Conference on Formal Ontology in Information Systems*. Amsterdam: IOS Press; (1998), 3–15.

- [4] Giandomenico P, Stefano B, Alessandro O et al. Information and organization in public health institutes: an ontology-based modeling of the entities in the reception analysis report phases. *Journal of Biomedical Semantics* (2016), 7:51.
- [5] Unified Medical Language System (UMLS) [EB/OL]. <https://www.nlm.nih.gov/research/umls/quickstart.html>, 2016-4-20/2017-4-10.
- [6] Open Biomedical Ontologies [EB/OL]. <https://sourceforge.net/projects/obo/>, 2017-3-7/2017-4-10.
- [7] OpenGALEN[EB/OL]. <http://www.opengalen.org>, 2012-3-6/2017-4-10.
- [8] Alexandre G, Jacques B, Suzanne P et al. An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders. *MEDINFO 2015: eHealth-enabled Health*. doi:10.3233/978-1-61499-564-7-275.
- [9] Uschold M, Gruninger M. Ontologies: Principles, methods and applications [J]. *Knowledge Engineering Review* 1996, 11(2): 14–17.
- [10] Gruninger M, Fox M S. Methodology for the design and evaluation of ontologies [EB/OL]. 2011-08-09. <http://stl.mic.utoronto.ca/publications/gruniger-ijcai95.pdf>.
- [11] Fernandez M, Gomez-Perez A, Juristo N. Methontology: From ontological art towards ontological engineering[C]. *Proceedings of the AAAI97 Spring Symposium on Ontological Engineering*. Stanford University, March 24–26th, 1997: 33 – 40.
- [12] Noy N F, McGuinness D L. Ontology development 101: A guide to creating your first ontology [EB/OL]. [2010-12-20]. <http://www.sfu.ca/dmarques/federation/pdf/Noy-Ontology101.pdf>. 2001.

Address for correspondence

Hu Hongpu (bubbleguo@163.com)

Interoperability of Medication Classification Systems: Lessons Learned Mapping Established Pharmacologic Classes (EPCs) to SNOMED CT

Scott D Nelson, PharmD, MS;^a Jaqui Parker, RPh;^b Robert Lario, MSE, MBA;^c Rainer Winnenburg, PhD;^d Mark S. Erlbaum, MD, MS;^b Michael J. Lincoln, MD, MS;^{d,e} Olivier Bodenreider, MD, PhD^f

^aDepartment of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; ^bApelon Inc, Harford, CT, USA;

^cDepartment of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA; ^dStanford Center for Biomedical Informatics Research, Stanford, CA, USA; ^eUS Department of Veterans Affairs, Spokane, WA, USA; ^fNational Library of Medicine, Bethesda, MD, USA

Abstract

Interoperability among medication classification systems is known to be limited. We investigated the mapping of the Established Pharmacologic Classes (EPCs) to SNOMED CT. We compared lexical and instance-based methods to an expert-reviewed reference standard to evaluate contributions of these methods. Of the 543 EPCs, 284 had an equivalent SNOMED CT class, 205 were more specific, and 54 could not be mapped. Precision, recall, and F1 score were 0.416, 0.620, and 0.498 for lexical mapping and 0.616, 0.504, and 0.554 for instance-based mapping. Each automatic method has strengths, weaknesses, and unique contributions in mapping between medication classification systems. In our experience, it was beneficial to consider the mapping provided by both automated methods for identifying potential matches, gaps, inconsistencies, and opportunities for quality improvement between classifications. However, manual review by subject matter experts is still needed to select the most relevant mappings.

Keywords: Topical; Pharmaceutical Databases

Introduction

Medication terminologies and ontologies commonly categorize medications by similar properties such as therapeutic intent (i.e., muscle relaxants or analgesics), chemical structure (i.e., sulfonyleureas or tetracyclines), mechanism of action (i.e., proton pump inhibitors or beta-adrenergic blockers), or sometimes combinations of the above (i.e., tricyclic antidepressants or amphetamine anorectics). Sets of drugs that share the same property used as a classification criterion (e.g., therapeutic intent) are generally referred to as medication classes (or simply classes in the context of this work). Interoperability among medication classification systems is known to be limited [4], yet it is important for clinical decision support (CDS), allergy checking, translational research, and organizing medication lists [3]. Our objective is to investigate the mapping of the U.S. Food and Drug Administration (FDA) established pharmacologic class (EPC) concepts to the SNOMED CT Substance hierarchy. More specifically, we provide an evaluation of ontology matching techniques and we describe lessons learned mapping medication classifications.

Established pharmacologic classes (EPCs)

In January 2006, the FDA established requirements for prescribing information for pharmaceutical products [2]. The labeling revisions provided additional information and established the structured product label (SPL) format for

prescription medication labeling in order to make it easier for health care professionals to access, read, and use prescription medication information. Part of the labeling revisions requires that the statement “(Drug) is an (EPC) indicated for (indication(s))” appear under the *Indications and Usage* section [1]. The EPC membership is determined by the FDA to classify the medications into medication classes to which the active ingredient belongs; for example, albuterol is a beta2-adrenergic agonist. However, unlike many other medication classes, the EPCs are not organized into a hierarchy, despite the presence of logical groupings in the EPCs. The absence of an EPC hierarchy makes the use of EPCs difficult for accessing and using prescription medication information, and may limit the use of SPLs in clinical decision support [12]. In contrast, SNOMED CT does have a robust medication class hierarchy and could be used to help organize the EPCs.

Medication class representation in SNOMED CT

SNOMED CT is maintained and distributed by SNOMED International (London, UK). SNOMED CT includes clinical terms used in healthcare, among which are two medication hierarchies, namely the *Pharmaceutical/biologic Product* and *Substance* hierarchies. However, for this study, we focused on the *Substance* hierarchy. Medications in SNOMED CT can belong to multiple medication classes, as can the medication classes themselves. For example, Figure 1 shows the medication classifications in the *Substance* hierarchy for albuterol are listed as a *Respiratory sympathomimetic agent*, *Selective beta-2 adrenoceptor stimulant*, and *Ethanolamine*.

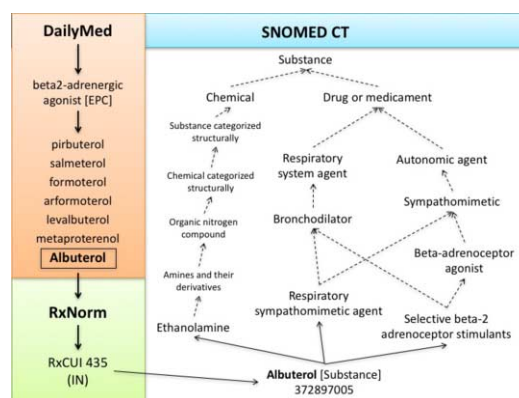


Figure 1 – Matching medication classes between EPCs and SNOMED CT, albuterol example. Note: Asserted membership shown with solid arrows and inferred membership shown with dashed arrows. See text for detailed explanation.

Ontology matching

Medication classes are present in most medication ontologies. Comparing and matching ontologies is not a new concept, and can be performed through various techniques [6]. Due to the time and labor-intensive nature of manual matching between ontologies, automatic techniques, such as lexical and instance-based matching have been developed [8; 9; 11]. Lexical matching compares medication classes based on their names (such as *Proton Pump Inhibitor [EPC]* matching with *Proton pump inhibitor [Substance]*) and is probably the most common technique for assisting manual matching. However, some authors have suggested that lexical matching may not be appropriate for comparing medication classes [7; 11]. On the other hand, instance-based matching compares the overlap of medication class members (instances) from one medication class to those of another, which may better represent the intended meaning of the medication class than using the class name. Winnenburger *et al.* used lexical and instance-based matching techniques as tools for medication class ontology matching [11], and Mortensen *et al.* used instance-based matching to compare medication classes between ontologies as a method of quality assurance to identify medication classes in need of review and/or updating [8]. However, these prior studies did not evaluate these ontology matching techniques against an expert-reviewed reference standard or provide guidance on which threshold for medication class match significance for instance-based matching.

The specific contribution of this study is to evaluate automated mapping techniques against a reference standard with application to the mapping of EPCs to SNOMED CT medication classes. Moreover, we conducted a sensitivity study to determine optimal thresholds for the instance-based techniques suggested by Winnenburger *et al.*

Methods

Data sources

We used the DailyMed index file (February 2014) to develop the list of EPCs with their corresponding medication unique identifiers (UNIs), and SNOMED CT (March 2014 release) for the *Substance* hierarchy. We used RxNorm (March 2014 release) to map medications between the EPCs and SNOMED CT using the RxNorm concept unique identifier (RxCUI) for the medication active ingredient (IN). RxNorm represents medications as ingredients (INs), precise ingredients (PINs), and multiple ingredients (MINs). We excluded MINs from the analysis because they may be represented as individual ingredients in the hierarchies and are inconsistently represented in medication classification systems.[10] For example, Combivent® (MIN RxCUI:214199) does not exist in any medication classes, but the ingredients ipratropium (an anticholinergic agent) and albuterol (a beta2-adrenergic agonist) do belong to medication classes. The PINs typically represent salt forms and esters of INs, so we normalized the medications to IN RxCUIs before analysis. For example, albuterol maps to an IN (RxCUI:435) whereas albuterol sulfate maps to a PIN (RxCUI:142153), yet both would be considered the same active ingredient, albuterol, so we normalized them to the IN (RxCUI:435).

Identifying medications and medication classes

As shown in Figure 1, we first obtained a list of EPCs with their respective medications from DailyMed. In SNOMED CT, medications are mixed in with the medication classes at

varying levels in the hierarchy, which can make separating the medications from the medication classes challenging. Therefore, for each medication (such as albuterol), we used RxNorm to identify the IN RxCUI (435 for albuterol), and then used RxNorm to map albuterol to the corresponding SNOMED CT concept identifier (372897005 for *Substance*), we then walked up the SNOMED CT hierarchies to identify medication classes with albuterol listed as a medication.

For example, albuterol has asserted membership to the medication classes *Respiratory sympathomimetic agent*, *Selective beta-2 adrenoceptor stimulant*, and *Ethanolamine* in the *Substance* hierarchy. We would then infer albuterol membership in ancestor medication classes through transitive closure. For example, since *Selective beta-2 adrenoceptor stimulant* is a subclass of *Beta-adrenoceptor agonist*, we would, therefore, infer that albuterol was also a member of *Beta-adrenoceptor agonist*, which would continue up the hierarchy to a *Sympathomimetic* and *Autonomic agent*. Figure 1 shows the asserted relationships with solid arrows and inferred relationships using dashed arrows. To simplify the analysis, we excluded very broad, top-level classes, such as *Drug allergen (Substance)*, *Chemical (Substance)*, and *Drug or medicament (Substance)*, which would not provide meaningful alignment.

Mapping medication classes from EPC to SNOMED CT

Lexical matching techniques

Lexical matching compares medication classes based on their name. Some lexical matching techniques, such as those used in this project, include exact, normalized, and approximate text matching applied to main terms and synonyms. For example, *beta2-adrenergic agonist [EPC]* would match (partially) with *Selective beta-2 adrenoceptor stimulant* though synonymy and normalization despite differences in hyphenation (ie, beta2 vs beta-2) or use of *adrenergic agonist* vs *adrenoceptor stimulant*. For lexical matching, we utilized the matching and searching algorithms in Termworks (Apelon Inc., 2013, Hartford, CT). Termworks would normalize the medication class names by stemming (ie. matching based on the root of the word, such as “stimul” from “stimulant, stimulants, stimulating, stimulator”) and then provide the top lexical match based on an internal scoring system, giving priority to exact, normalized, then approximate matches, in decreasing order. Termworks used synonyms from the target terminology, SNOMED CT.

Instance-based matching techniques

Instance-based matching techniques compare classes based on the class members they share, in our case medications. Here, we adopted the framework by Winnenburger *et al* [11]. In summary, EPC concepts were compared to each class in SNOMED CT in a pairwise manner if the classes shared at least one active ingredient. We calculated medication class similarity using the equivalence score (ES) between two classes as the modified Jaccard coefficient to account for small sample size in medication classes [11]. The ES gives a score from 0 to almost 1, with 0 meaning no overlap and higher scores representing greater similarity (overlap) of medication classes. For example, *beta2-adrenergic agonist [EPC]* has 7 ingredients, of which, all 7 overlap with the 7 ingredients in *Selective beta-2 adrenoceptor stimulant (substance)*, giving an ES of 0.941. The Winnenburger framework also supports the identification of inclusion relations between classes through an inclusion score. However, in the present investigation, we ignored the inclusion score and focused on equivalence relations between

classes. Analysis was completed using STATA 13 (StataCorp. 2013. College Station, TX).

Developing a reference standard

The reference standard was developed to contain mappings from the EPCs to SNOMED CT *Substance* hierarchy. To summarize, two pharmacist informaticists (SN and JP) served as subject-matter experts (SMEs). First, SMEs elicited the meaning of the EPC name and mapped it to an equivalent class in SNOMED CT; then, if there was no equivalent class, the EPC was mapped to a related class using *child_of* relationships. If both attempts failed, we concluded that no mapping could be established. During the mapping exercise, the SMEs independently reviewed and rated the top lexical EPC to SNOMED CT pairs from Termworks, and all instance-based EPC to SNOMED CT pairs with an ES ≥ 0.2 . An ES ≥ 0.2 was qualitatively chosen to provide a limited number of pairs for manual review that had a reasonable overlap. Priority was given to equivalent relationships, if no equivalent relationship was found, then the *child_of* relationships were used. If the pair was not equivalent, SMEs sought the most proximal SNOMED CT concept providing a more general, yet true representation of the EPC concept using *child_of* relationships. The SME ratings agreed 90.1% of the time with an average weighted kappa of 0.736, and discrepancies were resolved by discussion and consensus. Next, the ratings based on the automated methods were combined and the SMEs used clinical knowledge to review, investigate, and correct the mapping results, discrepancies, and unmatched EPCs. The reference standard underwent various quality assurance and verification reviews, including a full review of all mappings. The final mapping set was reviewed, and discrepancies reached consensus by an expanded group of physicians and FDA pharmacists.

Evaluating matching techniques against the reference standard

We used descriptive statistics to assess the contributions of each automated matching technique, along with their precision, recall, and F1 score (harmonic mean of precision and recall) compared to the reference standard. We also evaluated if the final equivalent pairs were predicted using lexical, instance-based matching, or both. To compare the lexical and instance-based matching, we limited the analysis to the top lexical match and top instance-based match using only equivalent pairs with the highest ES for each EPC. We then conducted sensitivity analysis across a range of ES thresholds for instance-based matching to determine which threshold provided the best F1 score for finding equivalent pairs.

Of note, in order to maximize the number of suggested matches (recall), we did not use the inclusion score to filter out those matches with a high ES that could also correspond to *child_of* relationships.

Since instance-based matching is dependent on shared instances, we removed EPCs that did not have medications matched to SNOMED CT. After statistical analysis, we reviewed valid, erroneous, and missing mappings, between lexical and instance-based matching techniques, compared to the reference standard.

Results

We identified 543 EPCs at the time of this study. There were 66 EPCs that were empty (i.e., no medications listed, such as *Acetaminophen [EPC]*, *Adrenergic Decongestant [EPC]*, and *alpha-Adrenergic and beta-Adrenergic Blocker [EPC]*), or did

not have medications that could be mapped to SNOMED CT using RxNorm (such as radioactive therapeutic agents like Iodine ion I-131 and various allergenic extracts). We identified 2,963 *Substance* classes that shared at least one medication with the EPCs.

Manually established reference standard

About half (284) of the 543 EPCs had an equivalent class in SNOMED CT, while most of the remaining (205) had a *child_of* relationship. Fifty-four EPCs could not be mapped to SNOMED CT, such as *sodium-glucose cotransporter 2 inhibitor [EPC]* (new therapeutic target not in SNOMED CT at the time of analysis), *calculi dissolution agent [EPC]* (mix of various medications with a common therapeutic intent, but not classified by therapeutic intent in SNOMED CT), and *potassium channel opener [EPC]* (somewhat vague mechanism of action for a specific medication).

Comparison to the reference standard

Using only the top match for each EPC, lexical matching identified 526 potential matches, whereas instance-based matching identified 282 potential matches.

Optimal threshold for the equivalence score (ES)

Sensitivity analysis across a range of ES thresholds showed that an ES ≥ 0.3 was the optimal threshold, maximizing precision and recall. Table 1 shows the performance of instance-based matching for identifying equivalent pairs with an ES ≥ 0.3 .

Table 1 – Performance of automatic methods from EPC to SNOMED CT for equivalent pairs (ES ≥ 0.3)

Method	Equivalent pairs	Precision	Recall	F1 score
Reference	284	-	-	-
Lexical	176	0.416	0.620	0.498
Instance-based	143	0.616	0.504	0.554

Contributions of each technique for equivalent mappings

As shown in Table 1, both lexical and instance-based matching had low-performance overall (F1 score was 0.498 and 0.554, respectively). Lexical matching had better recall, but lower precision than instance-based matching, which is not surprising, because the lexical matching produced more potential matches than the instance-based matching did.

Out of all 284 equivalent EPC to SNOMED CT pairs, there were 98 identified correctly only by lexical matching, 65 only by instance-based matching (ES ≥ 0.3), 78 identified by both, and 43 identified by manual review only.

Examples and failure analysis

Lexical matching techniques

There were cases where instance-based matching was able to identify a correct match, but lexical matching was not, and in some cases, lexical matching recommended a contradictory class as equivalent in attempts to maximize recall with approximate matches. For example, lexical matching suggested *Androgen Receptor Inhibitor [EPC]* be mapped to *Androgen receptor (substance)*; yet, instance-based matching correctly identified *Synthetic antiandrogen (substance)*.

Lexical mappings typically failed to identify matches due to the use of synonyms and ambiguous meaning of class name or differences in classification type, such as between structural and functional groupings or mixing structural and functional classifications. The lexical matching tended to suggest mappings to terms other than medication classes, (i.e.,

receptors, antigens, lab test ingredients), such as *Acetylcholine Release Inhibitor [EPC]* and *Acetylcholine (substance)*. In contrast, instance-based matched did not suggest pairs other than medication classes. Another limitation of lexical matching was that many of the suggested pairs had similar names but different meanings, which caused look-alike/sound-alike type errors for SMEs reviewing hundreds of pairs, such as *Mood Stabilizer [EPC]* and *Mast cell stabilizer (substance)*. Additionally, lexical matches were typically correct or incorrect, with very few "close" matches, whereas instance-based matching presented many close matches that pointed the mapper to the correct part of the SNOMED CT tree.

Instance-based matching techniques

Instance-based matching is intrinsically dependent on the classes sharing instances. Potential errors in the instance-based matching were mainly due to small numbers in class size. For example, *RANK Ligand Inhibitor [EPC]* with the medication denosumab was mapped by the SMEs as a child of *Bone resorption inhibitor (substance)*; however, *Bone resorption inhibitor* does not include denosumab, instead, denosumab is a child of *Monoclonal antibody agent (substance)*.

Instance-based matching also had problems with the multiple levels of granularity of SNOMED CT. For example, multiple medication classes were considered equivalent due to a high ES, such as *xanthine oxidase inhibitor [EPC]* matching with *Anti-gout agent (Substance)* with an ES of 0.516, and *Xanthine oxidase inhibitor (Substance)* with an ES of 0.775. Both substance classes were potential matches, but *Anti-gout agent (substance)* is the parent of *Xanthine oxidase inhibitor (Substance)*. This also occurred with *benzodiazepine [EPC]* matching with 6 medication classes from SNOMED CT. To help sort this out in our automated analysis, we only used the highest ES scoring pair for each EPC.

Additionally, there were some classes that had high ES ratings but were considered to be child_of relationships due to how the medication class was named, such as the use of qualifiers like "analog" or "recombinant". For example, *Folate Analog [EPC]* was considered a child of *Folic acid (substance)*. Finally, since instance-based matching is dependent on the medication classes having common medications, it was unable to find matches for EPCs without active ingredients listed in them, such as *Adrenergic Decongestant [EPC]*, or classes where the instances could not be mapped to the SNOMED CT.

Discussion

Lessons learned

Overall, comparing and mapping medication classification systems is a very challenging task. Each classification system has its own way of grouping medications, each using different grouping criteria. Thus it is not possible to simply insert one classification system into another since each classification system has its own "world view", use case, and area of interest, which has resulted in the proliferation of many terminology standards. However, the automatic matching methods helped facilitate the mapping process, for both equivalence and partial mappings. The lexical and instance-based matching methods were each able to correctly identify about half of the 284 manually created equivalent medication pairs (176 and 143, respectively). However, used in combination, the two approaches were able to find 221 of the 284 pairs, and identify most of the child_of relationships, indicating that these approaches have their own individual strengths and weaknesses. Each method contributed specific results, though they may occasionally contradict, we found

that providing both results (using OR logic) provided the best compromise and mapping facilitation. Our results suggest that a combination of lexical and instance-based matching could provide improved automated matching suggestions.

Previous studies have not been able to determine an ES threshold for instance-based matching. From our sensitivity analysis, we concluded that an ES threshold of 0.3 or greater provided a good balance between precision, recall and the number of pairs for SMEs to review. Despite the lexical bias in the development of the reference standard, instance-based matching performed slightly better than lexical matching in identifying equivalent pairs and was also useful later in identifying some child_of relationships. The instance-based scores presented in this paper improve upon those proposed by the framework and provide a reference of expected scores; however, scores may vary depending on use case and need to be validated in different data sets. Additionally, due to the incomplete nature of the EPCs, these values likely represent the lower bound of performance for instance-based mapping.

Application of results

First, we found these automated mapping methods to be useful in supporting the development of a mapping between medication classification systems. We want to point out that the goal of these automated matching methods is not to identify 100% correctly the first time, but to help narrow down the thousands of potential matches to a more manageable list so that someone knowledgeable of both terminologies can review, validate, or correct the mappings. Therefore, we found that manual mapping and SME review are still needed, but that the use of lexical and/or instance-based matching facilitated the process and could point the mapper to the right section in the other ontology (particularly true for the instance-based method). These matching methods also reduced the amount of curation needed for the mapping and can be generalized to virtually any pair of medication classification systems (beyond EPCs and SNOMED CT).

Second, we also found these methods useful for quality assurance and validation. For example, both EPC concepts and SNOMED CT classes switched between grouping active ingredients by chemical structure and by therapeutic intent, which was variable within and between the two classification systems; however, we found that the instance-based method was able to inform the meaning of the class names, such as if the class was grouped using structural or functional properties. The variation of naming and grouping medications was especially apparent with the EPCs, such as the use of *antibacterial* versus *antimicrobial* classifications. For example, the EPC concept *Macrolide [EPC]* only has erythromycin listed in it, *Macrolide antibacterial [EPC]* has only fidaxomicin listed in it, and *Macrolide antimicrobial [EPC]* has azithromycin, clarithromycin, and erythromycin listed as instances, yet all these instances are antibacterial agents with a macrolide structure, while other non-antibacterial agents with a macrolide structure (ie. tacrolimus or sirolimus) are not included in the *Macrolide [EPC]* class. One would expect that the concept *Macrolide [EPC]* would contain all medications with a macrolide structure, not just erythromycin. These inconsistencies arise because there was previously a lack in standardization in how EPCs were assigned [4]. We also found apparent gaps in SNOMED CT related to agonist/antagonist classifications, such as *Dopamine-2 Receptor Antagonist [EPC]* and *Gamma-Aminobutyric Acid A Receptor Agonist [EPC]*. Inconsistencies and unused medication classes were reported to the FDA and SNOMED International to help improve the medications classifications. Since then, SNOMED International has been

working on revising their medication classes in the *Substance* hierarchy. The identification of these types of inconsistencies using lexical and/or instance-based matching may provide opportunities for review, validation, correction, and further classification development [8].

Thirdly, since conducting this project, the results of this work helped to inform the development of a new RxNorm tool from the National Library of Medicine named RxClass (<https://mor.nlm.nih.gov/RxClass/>) [5]. RxClass provides a graphical web interface and application programming interfaces (APIs) for exploring medication class structures from various sources, as well as their corresponding RxNorm ingredients. RxClass can handle the IN/PIN normalization and find similar classes (with equivalence and inclusion scores) using instance-based matching across the various classifications in RxNorm.

Limitations

This study only mapped EPCs to equivalent SNOMED CT medication classes, and may not meet all of the needs or use cases of others. As such, we did not use inclusion scores, which would have represented the EPC as a child_of or parent_of the SNOMED CT class. Additionally, for our use case, our reference standard was developed based on the medication class intention, not always on extension, and as such was biased in favor of lexical matching, as we did not consider class instances in the final mappings. The mappings between EPCs and SNOMED CT were found to be mainly useful for navigation purposes, and their utility in clinical decision support remains to be studied, especially as the mappings did not account for inferred relationships caused by mappings. EPC concepts that could not be mapped to SNOMED CT could have medication classes added to SNOMED CT using extensions.

Secondly, instance-based matching is dependent on classes sharing instances. There were many empty and incomplete EPC classes, such as *Narcotic Antitussive [EPC]* not having any instances, dextromethorphan not listed as an antitussive, and erythromycin as the only instance in *Macrolide [EPC]*. Additionally, this study was limited to active ingredients mapped in RxNorm with RxCUIs and SNOMED CT identifiers. Some active ingredients were excluded because they are outside of the scope for RxNorm, such as allergens, pollens, foods, and herbals. These results likely represent the lower bound of the instance-based mapping performance.

Conclusion

Comparing and mapping between medication classes is challenging task due to the different ways of classifying medications. Using lexical and instance-based matching, with manual review, we were able to map most of the EPCs to medication classes in SNOMED CT. Each method had its own unique strengths, weaknesses, and contributions. The use of instance-based matching in addition to lexical matching can help map or compare medication classes. The evaluation and comparison of ontologies is a complex process, and while these automated matching techniques can help, manual review by subject-matter experts is still needed for mapping between medication classification systems.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine (NLM).

During the project, Dr. Nelson was supported by the VA Advanced Fellowship Program in Medical Informatics of the Office of Academic Affiliations, Department of Veterans Affairs.

We would like to thank Paul C. Brown, PhD for his review and assistance in preparing this manuscript.

Disclaimer

The views, findings, and conclusions expressed in this report are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, the Food and Drug Administration, or the National Library of Medicine.

References

- [1] 21 CFR 201.57(a)(6), in, Federal Register, 2006.
- [2] Requirements on Content and Format of Labeling for Human Prescription Drug and Biological Products, in: H. Food and Drug Administration, ed., Federal Register, 2006.
- [3] O. Bodenreider, Biomedical ontologies in action: role in knowledge management, data integration and decision support, *Yearb Med Inform* (2008), 67-79.
- [4] O. Bodenreider and D.D. Fushman, Investigating drug classes in biomedical terminologies from the perspective of clinical decision support, *AMIA Annual Symposium Proceedings* **2010** (2010), 56-60.
- [5] O. Bodenreider, L.B. Peters, and T. Nguyen, RxClass-Navigating between Drug Classes and RxNorm Drugs, in: *ICBO*, 2014, pp. 106-107.
- [6] J. Euzenat and P. Shvaiko, *Ontology matching*, Springer, 2007.
- [7] T. Merabti, H. Abdoune, C. Letord, S. Sakji, M. Joubert, and S.J. Darmoni, Mapping the ATC classification to the UMLS metathesaurus: some pragmatic applications, *Studies in Health Technology and Informatics* **166** (2011), 206-213.
- [8] J. Mortensen and O. Bodenreider, Comparing Pharmacologic Classes in NDF-RT and SNOMED CT, in: *Proceedings of the Fourth International Symposium on Semantic Mining in Biomedicine*, 2010, pp. 119-124.
- [9] F. Mouglin, A. Burgun, and O. Bodenreider, Comparing drug-class membership in ATC and NDF-RT, in: *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, ACM, 2012, pp. 437-444.
- [10] H. Saitwal, D. Qing, S. Jones, E.V. Bernstam, C.G. Chute, and T.R. Johnson, Cross-terminology mapping challenges: a demonstration using medication terminological systems, *Journal of Biomedical Informatics* **45** (2012), 613-625.
- [11] R. Winnenburg and O. Bodenreider, A Framework for Assessing the Consistency of Pharmacological Classes across Sources, *Journal of Biomedical Semantics* (2014).
- [12] Q. Zhu, G. Jiang, and C.G. Chute, Profiling structured product labeling with NDF-RT and RxNorm, *J Biomed Semantics* **3** (2012), 16.

Address for correspondence

Scott D. Nelson, PharmD, MS
Assistant Professor, Department of Biomedical Informatics
Vanderbilt University Medical Center, Nashville, TN, USA
Scott.Nelson@Vanderbilt.edu

Interoperability of Disease Concepts in Clinical and Research Ontologies: Contrasting Coverage and Structure in the Disease Ontology and SNOMED CT

Satyajeet Raje, Olivier Bodenreider

U.S. National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Abstract

Objectives. To contrast the coverage of diseases between the Disease Ontology (DO) and SNOMED CT, and to compare the hierarchical structure of the two ontologies. **Methods.** We establish a reference list of mappings. We characterize unmapped concepts in DO semantically and structurally. Finally, we compare the hierarchical structure between the two ontologies. **Results.** Overall, 4478 (65%) the 6931 DO concepts are mapped to SNOMED CT. The cancer and neoplasm subtrees of DO account for many of the unmapped concepts. The most frequent differentiae in unmapped concepts include morphology (for cancers and neoplasms), specific subtypes (for rare genetic disorders), and anatomical subtypes. Unmapped concepts usually form subtrees, and less often correspond to isolated leaves or intermediary concepts. **Conclusion.** This detailed analysis of the gaps in coverage and structural differences between DO and SNOMED CT contributes to the interoperability between these two ontologies and will guide further validation of the mapping.

Keywords:

Biological Ontologies; Systematized Nomenclature of Medicine; Unified Medical Language System

Introduction

Different ontologies are used to represent disease concepts in biomedical research and in clinical settings. The Disease Ontology (DO) is widely used in the research community, especially in genomic and cancer research. SNOMED CT is primarily used in healthcare and clinical settings. Interoperability between these two important ontologies is critical for translational applications in biomedicine. For example, research findings about a disease (coded with DO) and clinical findings from EHR data about the same disease (coded with SNOMED CT) can be analyzed together only if the DO and SNOMED CT codes for the disease are mapped together.

In this paper, we investigate the coverage of disease concepts between DO and SNOMED CT. More specifically, we identify and characterize the concepts present in DO but not covered by SNOMED CT. We also analyze the differences in hierarchical structure between the two ontologies.

Background

Resources

Disease Ontology

The Disease Ontology (DO) [1] is part of the Open Biomedical Ontologies (OBO) [2] collection and is used in several research projects. The ontology is implemented using description logics (DL) and available in the Web Ontology Language (OWL) format. We worked with the August 2016 release of

the Disease Ontology. This version has 6931 active disease concepts. Some concepts in DO have explicit cross-references (represented by “obo:hasDbXref” relations) to concepts from SNOMED CT and other bio-ontologies. In this paper, we refer to these cross-references as “mappings”.

SNOMED CT

The Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) is the largest clinical terminology. We used the March 2016 release of SNOMED CT (US Edition), as this is the version cross-referenced by the August 2016 version of DO. SNOMED CT contains over 300,000 clinical concepts, with about 100,000 disease concepts. As for DO, SNOMED CT is developed using description logics. However, since SNOMED CT is distributed in a proprietary format, we converted it to OWL using the script provided as part of the release. We processed the OWL versions of DO and SNOMED CT using the Java OWL API.

UMLS

The Unified Medical Language System (UMLS) Metathesaurus [3], developed by the U.S. National Library of Medicine, provides mappings across concepts from various standard biomedical terminologies and ontologies, including SNOMED CT. However, the UMLS does not currently integrate the Disease Ontology. The UMLS provides a RESTful API to identify lexical matches among all the concepts from its sources. Each UMLS concept is linked to one of 15 Semantic Groups, including Disorders. Using UMLS allows us to leverage the rich synonymy across all its source vocabularies for mapping and the semantic characterization of its concepts for consistency checking. We used the 2016AA release of UMLS in this research as it contains the March 2016 release of SNOMED CT.

Related work

Kibbe et. al. [4] report on the overall coverage of the DO and its cross-references to other terminologies in their update on the Disease Ontology. In this study, we perform a deeper analysis of the cross-references to SNOMED CT specifically.

Previous work has investigated the coverage of concepts within specific subdomains of medicine [5]. It has been demonstrated that the UMLS semantics can be exploited successfully for mapping across vocabularies [6]. In previous work from our group, Dhombres et. al. [7] evaluated the coverage of phenotypes across the Human Phenotype Ontology (HPO) [8] and SNOMED CT. Fung et. al. [9] assessed coverage of rare diseases in ICD and SNOMED CT. In this paper, we use similar techniques for assessment of disease concepts between DO and SNOMED CT.

To the best of our knowledge, this is the first attempt to contrast the coverage and structure between DO and SNOMED CT.

Methods

Our approach to investigating the coverage and organization of disease concepts in DO and SNOMED CT can be summarized as follows. We first establish a reference list of mappings of DO concepts to SNOMED CT. We characterize the unmapped DO concepts semantically and structurally. Finally, we compare the hierarchical structure between the two ontologies.

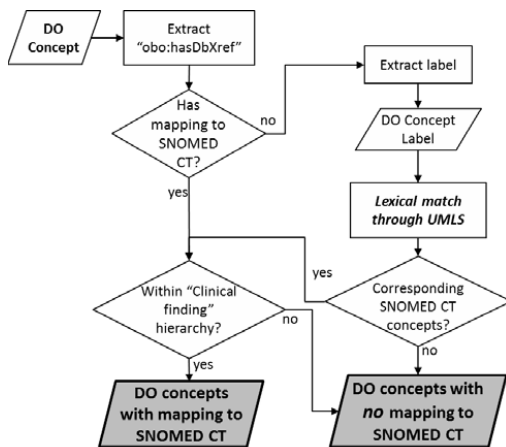


Figure 1- Methodology for establishing a reference list of mappings

Establishing a reference list of mappings

As shown in Figure 1, to establish a reference list of mappings, we start by updating the mappings provided by DO, from which we filter out semantically inconsistent mappings. We identify additional mappings lexically using the UMLS.

Updating and filtering mappings provided by DO

From the mappings (“cross-references”) to SNOMED CT provided by DO, we remove those mappings to retired SNOMED CT concepts and resolve the mappings to “moved” (remapped) concepts in SNOMED CT. Because DO concepts are expected to represent diseases, we consider semantically inconsistent and filter out those mappings to concepts outside the “Clinical finding” hierarchy of SNOMED CT, which contains all diseases, disorders and findings.

Finding additional lexical mappings

We leverage the UMLS in an attempt to identify lexical mappings for those DO concepts without any mappings to the “Clinical finding” hierarchy of SNOMED CT. More specifically, we first extract the labels for each DO concept, including preferred terms and synonyms. We take advantage of the rich set of synonyms provided by the UMLS to map these terms to UMLS concepts, using exact or normalized string matches. Finally, as we did for the mappings provided by DO, we select semantically consistent lexical mappings by restricting the mappings to the “Clinical finding” hierarchy of SNOMED CT.

All semantically consistent mappings (from DO cross-references or obtained lexically) form the reference list of mappings used in the rest of this investigation. All other DO concepts are considered unmapped.

Characterizing unmapped DO concepts

We characterize the unmapped DO concepts semantically and structurally, and analyze the differentia(e) between unmapped concepts and their parent(s).

Semantically. To identify whether coverage is better for some types of diseases than others, we compute the distribution of mapped and unmapped DO concepts with respect to the top-level subtrees of DO.

Structurally. To investigate whether unmapped DO concepts are isolated unmapped leaf concepts, subtrees of unmapped concepts, or unmapped intermediary concepts, we cluster them into groups of hierarchically related concepts (“connected components” in graph theory parlance).

Differentiae. Moreover, for isolated unmapped leaf concepts and subtrees of unmapped concepts, one author (OB) performed a manual review to identify the differentia(e) between each unmapped DO concept and its immediate parent concept(s). For example, the unmapped DO concept “ovarian germ cell teratoma” differs from its parent concept “ovarian germ cell cancer” by the addition of a morphology distinction (teratoma is a morphologic type of cancer).

Comparing hierarchical organization between DO and SNOMED CT

DO and SNOMED CT both represent disease concepts and share important classificatory principles (e.g., by localization, by etiology, by morphologic type). Therefore, we expect their hierarchical organization to be similar. In other words, we expect that most hierarchical relations present in ontology will also be present in the other. And we do not expect that two hierarchically related concepts in one ontology will have no hierarchical relation or will be considered the same concept in the other ontology.

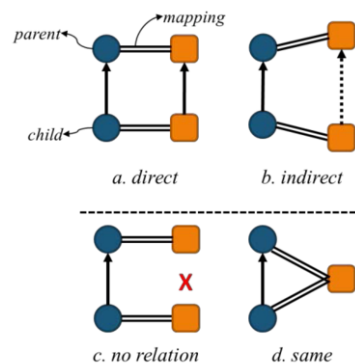


Figure 2- Types of possible hierarchical relations between corresponding mapped concepts

In practice, we take each pair of hierarchically related concepts (direct parent-child pair) present in one ontology and examine the relations between the two concepts in the other ontology. As illustrated in Figure 2, we consider 4 patterns of hierarchical relations across ontologies.

1. The two concepts are in the same direct parent-child relation in both ontologies.
2. The hierarchical relation is direct in one ontology but indirect in the other. The two ontologies are consistent,

but the ontology with the indirect relation is finer grained than the other.

3. The hierarchical relation present in one ontology is missing from the other. The two ontologies are inconsistent.
4. The two hierarchically related concepts in one ontology map to the same concept in the other ontology. The two ontologies are inconsistent.

To make this comparison possible, we restrict it to pairs of concepts in which both concepts have a reference mapping to the other ontology. Since there may exist multiple mappings for an individual parent or child concept, we compare all possible parent-child pairs. We apply this method in both directions (i.e., both from DO to SNOMED CT concepts and from SNOMED CT to DO).

Results

Establishing a reference list of mappings

Updating and filtering mappings provided by DO

There were 12,470 mappings to SNOMED CT provided by DO as cross-references. We removed 224 mappings to retired SNOMED CT concepts and resolved 6552 mappings to moved concepts. A total 4195 DO concepts have one or more mappings to SNOMED CT (involving 8352 mappings).

Of the 8352 mappings, we filtered out 1184 semantically inconsistent mappings. After this filtering step, 336 DO concepts were left with no mapping. In most cases, a disease concept from DO was mapped to a concept in the “Morphologic abnormality” hierarchy of SNOMED CT. For example, “mixed cell type cancer” [DOID:154] mapped to “Mixed tumor, malignant (morphologic abnormality)” [SCTID:8145008].

Of the 3859 DO concepts with at least one semantically consistent mapping, 3334 had only semantically consistent mappings, while 525 had at least one inconsistent mapping.

Finally, of the 3334 DO concepts with only semantically consistent mappings, 1950 concepts had a single mapping (e.g., “Cycloplegia” [DOID:10033] mapped to “Cycloplegia (disorder)” [SCTID:68158006] only), while 1384 had multiple semantically consistent mappings. Of these 1384 concepts, 110 had a mapping to concepts in both the “Disease” and the “Clinical Findings” hierarchies, usually to a disease its associated finding. For example, “Mevalonic aciduria” [DOID:0050452] is mapped to “Mevalonic aciduria (disorder)” [SCTID:124327008] and “Hyperimmunoglobulin D with periodic fever (finding)” [SCTID:234538002].

Overall, as shown in Figure 3, of the 6931 disease concepts in DO, 3859 (56%) were mapped to SNOMED CT through at least one semantically consistent mapping provided by DO, 336 (5%) had only semantically inconsistent mappings, and 2736 (39%) were unmapped.

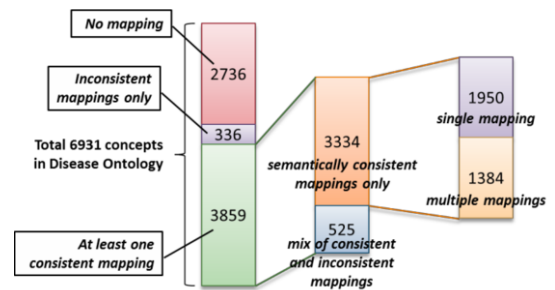


Figure 3- Breakdown of explicit mappings of the DO concepts

Finding Additional Lexical Mappings

Leveraging lexical mapping through the UMLS, we identified a mapping to SNOMED CT for 619 (20%) of the 3072 DO concepts with no semantically consistent mapping.

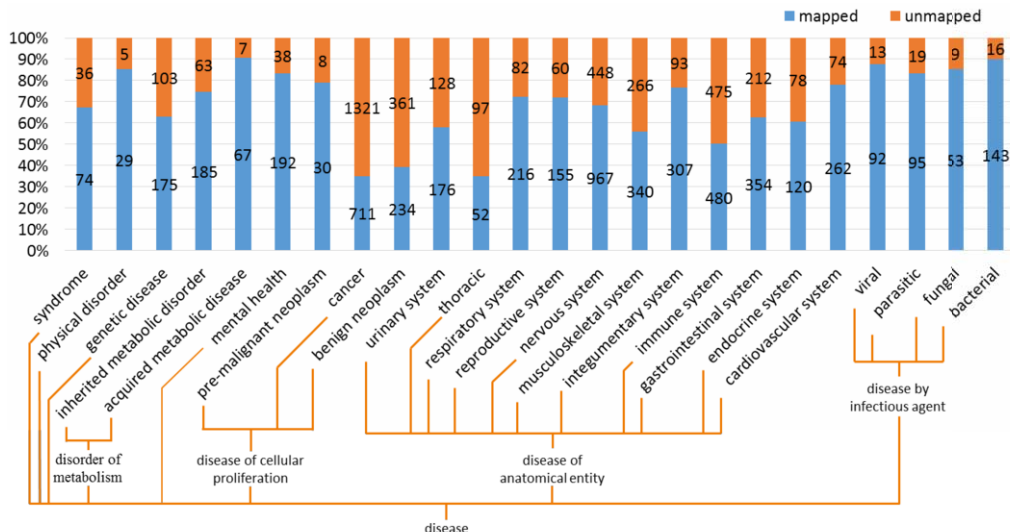


Figure 4- Distribution of mapped versus unmapped concepts within top level subtrees of the Disease Ontology. Numbers indicate the actual number of concepts in each subtree.

Overall, our reference list of mappings includes 7949 semantically consistent mappings covering 4478 (65%) the 6931 DO concepts, mapped to 6440 unique SNOMED CT concepts.

Characterizing unmapped concepts

Semantically. Of the 6931 DO concepts, 2453 (35%) remained unmapped to SNOMED CT. Figure 4 shows the distribution of mapped and unmapped concepts by the top-level subtrees of DO. Concepts belonging to multiple subtrees are counted multiple times. The top subtrees for unmapped concepts include cancers, neoplasms, diseases of the thoracic system, immune system diseases, and nervous system diseases. In contrast, very few of infectious diseases remain unmapped.

Structurally. The 2453 unmapped concepts can be grouped into 261 clusters of hierarchically related concepts (connected components). From a structural perspective, unmapped concepts fall under three possible categories: isolated unmapped leaf concepts, subtrees of unmapped concepts, and unmapped intermediary concepts.

We found 401 cases of **isolated unmapped leaf concepts**, namely 214 with a single parent and 187 with multiple parents. For example, “multiple mucosal neuroma” [DOID:5155] is the only unmapped child of “neuroma” [DOID:2001], which means that all of its siblings (e.g., “Neurilemmoma” [DOID:3192]), are mapped to SNOMED CT. Of note, 69 of these leaf concepts are the only child of their parent.

We found 1806 unmapped concepts in **subtrees of unmapped concepts**. These are clusters where none of the concepts is mapped, while they share a common mapped ancestor. For example, the subtree rooted at “chromosomal deletion syndrome” [DOID:0060388] contains 35 concepts, including “distal 10q deletion syndrome” [DOID:0060390] and “chromosome 15q11.2 deletion syndrome” [DOID: 0060393].

Finally, we found 246 cases of **unmapped intermediary concepts** (located between a mapped ancestor and a mapped descendant). 58 are intermediary “grouper” concepts in DO with all of its parents and children mapped to some SNOMED CT concept. 50 of these concepts sit between a single mapped parent and child. An example is “multifocal dystonia” [DOID:0050837]. This concept is unmapped to SNOMED CT, while its parent, “dystonia” [DOID:543], and child, “hemidystonia” [DOID:0050846], are mapped to SNOMED CT.

Table 1- Characterization of differentiae added by unmapped DO concepts

Type of differentiae	Count
Morphology (e.g. follicular dendritic cell sarcoma)	831
Morphology and anatomic site	520
Specific subtype (e.g. spinocerebellar ataxia type 1)	253
Anatomic site (e.g. intramuscular hemangioma)	147
Morphology and period of onset	61
Period of onset (e.g. pediatric osteosarcoma)	45
Chromosomal location and anomaly	45
Complex syndrome (e.g. agnathia-otocephaly complex)	42
Subtype	42
Organism (e.g. screw worm infectious disease)	30
Others	191

Differentiae. We examined all 2207 unmapped concepts that are isolated leaf or in subtrees, which represent the majority (90%) of the 2453 unmapped concepts, and analyzed how they differed from their parent concept(s). The most frequent differentiae, listed in Table 1 along with examples, include morphology (for cancers and neoplasms), specific subtypes (for rare genetic disorders), and anatomical subtypes. Of note, about a third of the unmapped concepts have more than one differentia. Typically these concepts have more than one parent. For example, “urethra adenocarcinoma” is a child of both “adenocarcinoma” (anatomical differentia) and “urethra cancer” (morphology differentia).

Comparing hierarchical organization between DO and SNOMED CT

We found 4233 direct parent-child pairs among the mapped DO concepts and 5772 among the mapped SNOMED CT concepts. After classifying each pair of hierarchically related concepts into the four patterns of hierarchical relations across ontologies presented earlier, we established the distribution of patterns shown in Table 2.

As mentioned earlier, in a given pair of hierarchically related concepts, each concept can be mapped to more than one concepts. Therefore, a pair of hierarchically related concepts can exhibit more than one pattern. To simplify the analysis, we distinguish between patterns indicative of semantic consistency (a and b) and patterns indicative of semantic inconsistency (c and d). Only about 30% of the hierarchical relations in DO and SNOMED CT are semantically consistent between the two ontologies (a/b only). The other hierarchical relations are either completely (c/d only) or partially inconsistent (a/b and c/d). This analysis reveals critical differences in hierarchical organization and concept orientation (i.e., whether two concepts correspond to the same entity) in the two ontologies.

Table 2- Characterization of the mapped parent-child concepts in comparison to the relation between their corresponding mapped concepts. The types are illustrated in Figure 2 above.

Mapping direction	Type (a) or (b) only	Type (c) or (d) only	(a or b) & (c or d)	Total pairs
DO to SNCT	1198 [28%]	1075 [25%]	1978 [48%]	4233
SNCT to DO	1842 [32%]	2792 [48%]	1138 [20%]	5772

Here are examples of relations for patterns. (The arrow, \rightarrow , represents the “child of” relation).

- For the DO relation “peliosis hepatis” [DOID:914] \rightarrow “hepatic vascular disease” [DOID:272], there is a direct corresponding relation in SNOMED CT, “Peliosis hepatis (disorder)” [SCTID:58008004] \rightarrow “Vascular disorder of liver (disorder)” [SCTID:235878005]. The two ontologies are perfectly aligned in this case.
- For the DO relation “diphtheritic cystitis” [DOID:13306] \rightarrow “cystitis” [DOID:1679], there is an indirect corresponding relation in SNOMED CT, “Diphtheritic cystitis (disorder)” [SCTID:48278001] \rightarrow “Bacterial cystitis (disorder)” \rightarrow “Infective cystitis (disorder)” \rightarrow “Cystitis (disorder)” [SCTID:38822007]. The two ontologies are semantically consistent, but SNOMED CT is finer-grained in this case as its two intermediary concepts are missing from DO.

- For the DO relation “portal hypertension” [DOID:10762] → “hepatic vascular disease” [DOID:272], there is no corresponding hierarchical relation in SNOMED CT between these two concepts. Instead, “Portal hypertension (disorder)” [SCTID:34742003] and “Vascular disorder of liver (disorder)” [SCTID:235878005] are in different parts of the “disorder of abdomen” hierarchy. In this case, the two ontologies are inconsistent.

Discussion

Pre- vs. post-coordination

Most biomedical terminologies, including DO, only consider pre-coordinated concepts. In other words, there is no built-in mechanism in DO to combine existing concepts to derive new concepts. As a result, only existing, pre-coordinated concepts are available to applications (e.g., for annotation purposes). In contrast, SNOMED CT supports post-coordination through a compositional grammar [10], which reflects semantic constraints expressed in the SNOMED CT concept model [11]. For this reason, SNOMED CT tends to adopt a parsimonious approach to pre-coordination, i.e., avoid pre-coordinating what can be expressed through post-coordination.

As shown in Table 1, the combination of differentiae morphology and anatomic site is the single most frequent combination. While pre-coordinated concepts are generally easier to use, the proliferation of pre-coordinated concepts may add unnecessary to the terminology.

Resolving multiple mappings

The mappings (cross-references) to SNOMED CT provided by DO frequently involve multiple SNOMED CT concepts (1-many mappings). Even after filtering out semantically inconsistent mappings (e.g., mapping of a disorder to a morphology concept), many multiple mappings remain.

In fact, in our reference mapping, 1384 (42%) of the 3334 DO concepts with mapping to SNOMED CT have multiple (semantically consistent) mappings to SNOMED CT. Of these, there are 110 concepts with a mapping to both a “Disease” concept and a “Clinical Findings” concept. In such cases, the “Disease” concept could be given precedence. The remaining 1274 DO concepts have multiple mappings within the same hierarchy. In this case, the structural analysis we performed can help guide the mapping. Mappings involved in semantically consistent patterns of hierarchical relations (namely type a and b) could be given precedence.

Conclusion

Overall, 4478 (65%) the 6931 DO concepts are mapped to SNOMED CT. The cancer and neoplasm subtrees of DO account for many of the unmapped concepts. The most frequent differentiae include morphology (for cancers and neoplasms), specific subtypes (for rare genetic disorders), and anatomical subtypes. Unmapped concepts usually form subtrees, and less often correspond to isolated leaf concepts or isolated intermediary concepts. This detailed analysis of the gaps in coverage and structural differences between DO and SNOMED CT contributes to the interoperability between these two ontologies and will guide further validation.

Acknowledgements

This work is supported in part by the Intramural Research Program of the U.S. National Library of Medicine (NLM) and an appointment to the NLM Research Participation Program administered by ORISE through an interagency agreement between the U.S. Dept. of Energy and the NLM.

References

- [1] L.M. Schriml, C. Arze, S. Nadendla, Y.W. Chang, M. Mazaitis, V. Felix, G. Feng, and W.A. Kibbe, Disease Ontology: a backbone for disease semantic integration, *Nucleic Acids Res* **40** (2012), D940-946.
- [2] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R.H. Scheuermann, N. Shah, P.L. Whetzel, and S. Lewis, The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nat Biotech* **25** (2007), 1251-1255.
- [3] O. Bodenreider, The Unified Medical Language System (UMLS): integrating biomedical terminology, *Nucleic Acids Res* **32** (2004), D267-270.
- [4] W.A. Kibbe, C. Arze, V. Felix, E. Mitraka, E. Bolton, G. Fu, C.J. Mungall, J.X. Binder, J. Malone, D. Vasant, H. Parkinson, and L.M. Schriml, Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data, *Nucleic Acids Res* **43** (2015), D1071-1078.
- [5] C.G. Chute, S.P. Cohn, K.E. Campbell, D.E. Oliver, and J.R. Campbell, The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures, *Journal of the American Medical Informatics Association* **3** (1996), 224-233.
- [6] O. Bodenreider, S.J. Nelson, W.T. Hole, and H.F. Chang, Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies, *Proceedings of the AMLA Symposium* (1998), 815-819.
- [7] F. Dhombres, R. Winnenburg, J.T. Case, and O. Bodenreider, Extending the coverage of phenotypes in SNOMED CT through post-coordination, *Stud Health Technol Inform* **216** (2015), 795-799.
- [8] P.N. Robinson, S. Köhler, S. Bauer, D. Seelow, D. Horn, and S. Mundlos, The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease, *The American Journal of Human Genetics* **83** (2008), 610-615.
- [9] K.W. Fung, R. Richesson, and O. Bodenreider, Coverage of rare disease names in standard terminologies and implications for patients, providers, and research, *AMIA Annu Symp Proc* **2014** (2014), 564-572.
- [10] IHTSDO, SNOMED CT Compositional Grammar Specification and Guide, in, 2016.
- [11] IHTSDO, SNOMED CT Concept Model, in: *SNOMED CT Starter Guide*, 2016.

Address for correspondence

Olivier Bodenreider, MD, PhD
8600 Rockville Pike, 38A/09S904, Bethesda, MD 20894 Phone Number: (301) 827-4982
E-mail: olivier.bodenreider@nih.gov

From Patient Discharge Summaries to an Ontology for Psychiatry

Marion Richard^{a,d}, Xavier Aimé^{a,d}, Marie-Christine Jaulent^a, Marie-Odile Krebs^{b,d}, Jean Charlet^{a,c,d}

^aINSERM, U1142, LIMICS, F-75006, Paris, France;

Sorbonne Universités, UPMC Univ Paris 06, UMR S 1142, LIMICS, F-75006, Paris, France;
Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR S 1142), F-93430, Villetaneuse, France.

^bLaboratoire de Pathophysiologie des Troubles Psychiatriques, Centre Hosp. Sainte-Anne, Paris, France

^cAssistance publique - Hôpitaux de Paris, France

^dGDR Psychiatrie CNRS 3557, Paris, France.

Abstract

Psychiatry aims at detecting symptoms, providing diagnoses and treating mental disorders. We developed ONTOPSYCHIA, an ontology for psychiatry in three modules: social and environmental factors of mental disorders, mental disorders, and treatments. The use of ONTOPSYCHIA, associated with dedicated tools, will facilitate semantic research in Patient Discharge Summaries (PDS). To develop the first module of the ontology we propose a PDS text analysis in order to explicit psychiatry concepts. We decided to set aside classifications during the construction of the module, to focus only on the information contained in PDS (bottom-up approach) and to return to domain classifications solely for the enrichment phase (top-down approach). Then, we focused our work on the development of the LOVMI methodology (Les Ontologies Validées par Méthode Interactive - Ontologies Validated by Interactive Method), which aims to provide a methodological framework to validate the structure and the semantic of an ontology.

Keywords:

Ontology; Psychiatry; Social Factors

Introduction

The descriptions of cases with mental disorders are various, heterogeneous and encompass many aspects of personal history, family history and many facets of dysfunctions and subjective symptoms. The large amount of information available in psychiatry stresses the need to improve our ability to detect, quantify the behavior and model the symptoms or dysfunctions associated with psychiatric disorders. Tackling the challenge of modeling information is one field of knowledge engineering. Ontologies are created for many purposes, including semantic researches, data retrieval and formal reasoning. The validation of the modeling is an important step in the process of ontology development. This step ensures the adequacy of the ontology and reduces maintenance costs.

We propose ONTOPSYCHIA, an ontology for psychiatry, divided into three modules: social and environmental factors of mental disorders, mental disorders and treatments. The use of ONTOPSYCHIA, associated with dedicated tools, will allow to perform semantic research in Patient Discharge Summaries (PDS), to represent co-morbidity, to index PDS for the constitution of cohorts, and to identify resistant patient profiles. We assume that this approach can help reach a consensus on descriptive categories of mental disorders as

well as a model to correlate social context, event and mental health. As part of this work, we developed a method for structural and semantic validation of ontologies. This method, LOVMI, relies on existing tools and an interactive approach with the experts of the modeled domain. In this paper, we present the development process of the module on social and environmental factors of mental disorders, and its validation with LOVMI methodology.

We conducted a first review of the literature in the domain of classifications and ontologies in psychiatry. We found that reference classifications in psychiatry (such as the ICD, the DSM or the RDoC project) are barely used in knowledge modeling. Furthermore, ontologies that model these classifications, such as the Mental Disease Ontology to describe and categorize mental disorders [1, 2], are usually not available. As regards these, and ontologies dealing with social issues such as SNOMED, much remains untapped in the field of mental illness. Therefore, we note a gap in the formal representation of psychiatric disorders and social factors impacting these diseases. The second part of our literature review confronted us with problems linked to ontology validation. Overall, the developed methods do not deal with all aspects of validation. Most of them only address the structural validation or the semantic validation. Finally, few offer easy-to-implement and semi-automatic solutions.

Methods

The distinctive features of our project have to do with the choice to model the environmental and social factors that can affect the patient's life and the choice of a modular approach. We chose a modular approach to deal with the amount of knowledge available in psychiatry. Each concept of ONTOPSYCHIA is denoted both by a preferred label in English and in French, and by alternative labels (synonymous terms, acronyms). The labels come from the list of extracted terms and from nomenclatures (DSM, ICD and SNOMED). To develop our ontology we followed the generic methodology illustrated by Charlet et al. [2] that combines a top-down and a bottom-up approach. This method allows access to terms that represent concepts in use. During the construction of the ontology, we focused only on information contained in PDS (top-down approach). Indeed, classifications modelling social or environmental factors as the SNOMED 3.5FR make some associations of terms that did not satisfy our expectations (e.g. "mendiant" (~ beggar) and "fainéant" (~ lazy person) are synonyms). We returned to domain classifications only to enrich our ontology (bottom-up approach).

Material

The corpus used in our work is composed of nearly 8,000 PDS, which represents a volume of around 5,804,000 words. These PDS are from the university hospital department of therapeutic and mental health of Sainte-Anne Hospital, the biggest psychiatric hospital in Paris. These PDS are semi standardized, in Word format and not anonymous. They cover a ten-year period. For several years, the diagnostic is provided at the end of the PDS with the CIM 10 code.

Terms extraction

We extracted specialized terms of our corpus with the term extractor YATEA, a tool used for identifying groups of words that can correspond to specialized terms in a text [3]. However, YATEA requires an input of corpus segmented in words and sentences tag with part of speech information using the annotator TREETAGGER. Unfortunately, the annotator TREETAGGER gave poor results for French: many words were not recognized and annotated as such. The rest of the treatments was impacted by this incorrect annotation. Therefore, we opted for the annotator MELT that offered a better annotation rate on our specialized corpus. Then, we developed a TREETAGGER converter for both format and tag in order to be able to use the corpus annotated by MELT with YATEA.

Domain conceptualization

To conceptualize the domain from the list of extracted terms previously, we followed recommendations of METHONTOLOGY [5]. Firstly, we built a glossary of validated terms into concepts with their definition(s) and all lexical equivalent(s) in natural language. Secondly, we built all subsumption relations, the taxonomy. Finally, we built relations between concepts. Throughout the development of the taxonomy, we considered the semantic commitment from Bachimont et al. [6]. This commitment involved the differential principle. This principle enables us to determine the significance of a concept according to its position in the tree (subsumption relation), by similarities and differences with neighbors (parents and siblings). It implies that no concept of an ontology can have two parents. The last step of the conceptualization is about ontology enrichment. This enrichment can be realized by alignment with other resources or by integration of new concepts from other resources.

LOVMI validation of the model: Ontology Validation with an Interactive Method

The method of validation is based on the criteria defined by Poveda-Villalón et al. [7]: human understanding, logical consistency, modeling issues, ontology language specification, real world representation and integration in semantic applications. They were established following a sound field study. Our method used a series of existing tools to validate the structure of the ontology. Interactive interviews with domain specialists were conducted to validate the semantic. LOVMI validation is performed in six steps: validation (1) of consistency, (2) of other structural aspects, (3) of labels, (4) of choices of label and (5) of semantic with experts and (6) of semantic in an application. These steps are neither linear nor cyclic. They may be carried out independently, in whole or in part. The user can skip a step to come back to it later.

Results

Material

In order to use our corpus, we had to anonymize the PDS. We used Medina [8] software (which has been developed following the HIPPA law). We anonymized seven ID including first and last name of patients, practitioners, and all persons quoted in the PDS, social number, address, date, names of hospital and phone number. This step has been made semi automatically. Medina helped us to annotate the named entities, but we had to check the validity of these annotations. Then we automatically removed all named entities annotated with Medina. All PDS have been checked by hand by authorized persons from the clinical department. This work required the equivalent of one person working full-time for a month.

Terms extraction

We extracted exactly 198,615 words. We removed all terms under a frequency of four in the corpus. Consequently, we dealt with an exhaustive list of 27,744 terms related to the psychiatry domain, to validate by hand. In this list, we removed numeral expressions, drug names, disorders and symptoms, leaving only concepts about social life or environmental events. One person, not specialized in the psychiatry domain, performed this task. This also explains the importance we gave to the validation of the modeling.

Domain conceptualization

Following METHONTOLOGY's recommendations, we first validated a list of concepts from our 27,744 terms related to the field of social and environmental factors of mental disorders. At this stage, we validated about 1,100 terms into concepts with their definitions and lexical equivalent(s) in natural language. Secondly, we started to develop the taxonomy. We divided the modeled domain in four major classes that never changed despite adjustments within these four classes: (1) Attribute – adjectives describing more precisely the concepts; (2) Concept of social life – concepts to organize all concepts extracted from the corpus and dealing with social aspects (e.g. education, social situation or civil situation); (3) Human being – to represent all human entities; and (4) Groups – to formalize the primary considered group, such as company, institution, service or clinic. Thirdly, we built relations between concepts. This taxonomy of relations evolved a lot from the first version of our ontology. At the beginning of the development, we had 12 relations only, whereas now we can count 199 relations between concepts.

Table 1 provides the concept number in each of the four major classes. The class "Concept about social life" is the main topic of the conceptualized domain (the social and environmental factors), and, consequently, it is the one containing more concepts. In addition, we can notice that the concepts under "human being" are highly connected. That can be explained by connections between individuals: 111 object properties are only describing relations between family members. Among these family members relations around 33 are connected with concepts about social life and others are connected with other human beings who are not family members (such as friends or boyfriends). We do not present the class reserved to attributes, as it is only a list of adjectives.

Table 1 – Some statistics on the social and environmental factors module. The classes numbers include the two parents concepts.

Social module	Classes	Object properties
OntoPsychia	1478 (252 with two parents)	199
Attribute	50	2
Concept about social life	1039	51
Human being	560	144
Group	81	2

In table 2, we notice that the subclass number is higher than the class number. This means that around 250 concepts possess two parents or more (only “Human Being” can have more than one parent, for example an “adoptive sister” is an “adoptive sibling” plus a “female individual” plus a “sister”). This does not conform to the differential principle. Basically, two types of explanations can be given to this remark: (1) the Family Health History Ontology (FHHO) which enriches the “Human Being” section contains more than 150 concepts with two parents; (2) some concepts of the ontology needed to be represented by two concepts. Our goal is to use this ontology in an annotation system. Consequently, it was easier, whenever possible, to model concepts with two parents rather than with a logical restriction.

Table 2 – Social and environmental factors module metrics

Number of classes	Number of properties	Maximum depth
1478	199	11
Maximum number of children	Average number of children	Classes with a single child
56	3	146
Classes with more than 25 children	Classes with no definition	Number of sub-classes
3	1434	252

Finally, we tried to align our model with the SNOMED-3.5 version française (French version). However, we quickly needed to move away from this modeling. Indeed, we identified some associations of terms that did not satisfy our expectations: for example *mendiant* (beggar) and *fainéant* (lazy person) are considered synonyms. Consequently, the module finally contains only 93 concepts aligned on the SNOMED-3.5VF. We decided to wait for an official French version of the SNOMED CT before trying a new alignment with our ontology. In order to enrich our model, we included the FHHO [9] that conceptualizes (1) members of family and relations between them, (2) and health states with diagnosis. Even though we were not interested in the second part of FHHO, we took all classes modeling family members. However, we re-organized them to match our modeling choices. In FHHO, all members are defined under a structuring concept (e.g. “Adopted Cousin” is an “Adopted Kin Relation”). In order to avoid the hierarchical system, we modified all structuring concepts into individual concepts (e.g. “Adopted Kin Relation” is in ONTOPSYCHIA the concept “Adoptive Individual”).

LOVMI validation of the model: Ontology Validation with an Interactive Method

In order to test our validation method, we used LOVMI to validate our module on social and environmental factors of mental disorders. We followed each step one by one.

Step 1 – Consistency validation

During the development of the ontology, we used HERMIT reasoner to regularly check the consistency of our model.

Step 2 – Validation of other structural aspects

Once the conceptualization ended, we used the on-line tool OOPS [7] to check human understanding, modeling problems and specification ontology language. We found several mistakes such as: two cases of the pitfall “creating unconnected ontology elements”; two cases of the pitfall “defining wrong inverse relationships”; 153 cases of the pitfall “missing domain or range in properties”; or the pitfall “using different naming conventions in the ontology” concerning the entire ontology.

Step 3 – Label validation

Then, we used SPARQL query language to check that all concepts had one unique preferred label in English and one unique preferred label in French. We counted three missing preferred label in French and 186 missing preferred label in English. This gap in English can be explained by a significant number of legal and educational concepts that cannot be translated in English. These concepts cover the French law or the French education system and not an international system that could be equivalent for everyone. Yet, to be consistent with our modeling choices, we decided to provide a lexical translation (so a preferred label in English) to a maximum of them. When it was absolutely not possible to find an appropriate lexical translation, we gave them the preferred label “No Equal Word”. On this day, we count 38 of these concepts.

During the validation of labels, we also found a total of about ten English preferred labels and two French preferred labels that were not single. Finally, we counted around 35 similar lexical preferred labels, which were not similar as language tag. For example, “discrimination” is the same lexical unit both in French and English, so it was appearing twice, once with the English tag and once with French tag.

Step 4 – Validation of the choice of labels

Once all these steps were completed, we only had to check whether our labeling choices were correct. We used the method from Aimé and Charlet [10]. Following this method, we found out that for 76 concepts (about 5% of the concepts) we had to change their preferred label according to the usage of this concept in the corpus. For example, some acronyms are more common than their developed form such as SAMU for “service d’aide médicale urgente” (“emergency service”), or some synonyms such as “emploi” (“job”) for “travail” (“work”).

Step 5 – Validation of semantic in interaction with experts

We implemented an interactive validation method in order to foster a common understanding of the ontology. This method is based on communication between domain players and ontologists who participated in the development of ONTOPSYCHIA. To involve the domain players, we reported on our project through meetings, presentation tools or presentation on the project impact assessment. Simultaneously, we made our ontology available for domain players via WEBPROTÉGÉ. Each player could visualize the conceptual taxonomy of the ontology (including axioms and relationships). Then, we asked them to meet in small groups to

enable active discussions on the modeling. In case of disagreement, the domain players were invited to raise their concerns and formulate solutions.

Validation sessions were held in groups of two (a clinical psychologist and a psychiatrist) and lasted about two hours. Each player had his own computer and therefore access to the ontology on WEBPROTÉGÉ. Each group had to work to interact on the same concepts to initiate discussions and debates. The conversations were recorded to allow the ontologist to keep a record of all interviews. Players were invited to comment on free text on WEBPROTÉGÉ, as a summary of the points raised during discussions on a concept or on a branch of concepts. This contributed to the interactions between the actors. Those not present during the validation session had access to discussions, could provide answers or participate. Once these recommendations were set, the ontologist did not provide further guidance to the actors. He interacted only with them to explicit modeling choices found ambiguous by the actors. For example, the concept “education” could refer to “public education” or “parenting education”, in such case it is important to verbally disambiguate the concept definition (even if this definition is already included in the ontology.)

To validate all modules on environmental and social factors that can affect the patient’s life, we organized four validation sessions. Each session was conducted by the ontologist who developed the model and by two practitioners (a clinical psychologist and a psychiatrist). It allowed us to validate the tree, but not the resulting conceptual relationships. We calculated that we can validate in average 164 concepts per hour. However, each concept or branch of concepts does not require the same validation time. In our modeling, we found that the concepts fall into two categories according to the denoted degree of subjectivity. The more subjective a concept is, the longer is the validation. The time specified here is based on recordings from the first group of the first validation phase.

The concepts not subject to interpretation concern the branches that model objective formal concepts, with little latitude for interpretation. For example, we can mention the concepts modeling “school education” as “educational establishment” or “schooling”. We counted 124 of these concepts for the first validation phase. They were validated very quickly, without asking questions. In total, we recorded 43 seconds of oral communication time and five written comments on all of these concepts.

The concepts subject to definition, contextual or personal interpretation concern the objective concepts, but subject to a personal, contextual or definitional interpretation. We have for example to define with domain players whether a “companion” is perceived differently from a “husband”, and therefore used differently in the domain language; or whether the term “intimate relationship” means in the context of the domain a “very close emotional relationship between two individuals” or a “sexual relation”. We counted 97 of these concepts validated in around 40 minutes, so we estimated an average validation time for definitional and contextual interpretation. Finally, the meaning of certain concepts can be perceived very differently from one individual to another. These concepts have led to deep conversations to reach a consensus around the modeling. Specifically, the concepts related to professional events such as “a dismissal” can be perceived negatively or positively (e.g. as relief, in the case of a person suffering from burn-out). We counted 13 of these concepts subject to personal interpretation in the first validation phase. Each of these concepts generated about 15 minutes of discussion and seven reviews.

We can draw several conclusions from these validation sessions. In these four groups, no dominant appeared; everyone was speaking from his own experience and professional skills. The full view of the conceptual hierarchy helped players to understand the meaning of concepts and their level of relevance in the ontology. This enabled them to understand the modeling strategies. In some cases, it also helped the players to validate entire branches of concepts. This allowed a considerable time saving. Finally, viewing the axioms that define the classes enabled players to identify gaps (e.g. for the ontologist, the concepts “home” and “accommodation” are residential places, as for the leaders it was essential to add the definition of the type of housing as “collective” or “individual”).

Step 6 - Validation in semantic application

The last step of the validation process consists of using the ontology in a dedicated application. In order to identify patient profiles, we need to annotate our corpus with concepts modeled in ONTOPSYCHIA. To that end, we designed a prototype of an annotation system to automatically find concepts in our corpus. We developed our system with UNITEX/GRAMLAB, “an open source, cross-platform, multilingual, lexicon and grammar based corpora processing suite. UNITEX is a corpora-processing system, based on automata-oriented technology. With this tool, you can handle electronic resources such as electronic dictionaries and grammars and apply them. You can work at the levels of morphology, lexicon and syntax”.

We tested our system with a sample of 20 PDS with a total of around 15,270 words. Firstly, we envisaged the corpus annotation by hand, followed by a comparison of the system’s results. However, we encountered difficulties in identifying social concepts when reading PDS, and only one person was available for this work. Consequently, we decided to change our method and we first annotated the PDS with the system. Then, the only corrector considered whether the annotations were correct and identified the missing annotations.

In table 3, the number of 724 total annotated concepts includes 706 annotated preferred labels, and, consequently, 18 annotated alternative labels. This result confirms again that our choice of preferred labels was relevant. Indeed, only a few alternative labels have been annotated. We also count 401 annotated concepts of the social branch, 238 annotated concepts of the human being branch and 67 annotated concepts of the group branch. Then, we observe a good rate concerning precision and recall. Specifically, we have 18.1% of missing annotations and 16.2% of annotations performed that are incorrect. However, we observe that the incorrect annotations are not caused by a poor labeling choice, but rather by faulty analysis of the context of appearance of the concept. For example, the concept “trouble” never appears alone in PDS and yet the system allowed its annotation without context (e.g. trouble are always defined according to what is troubled: “attention trouble”). Therefore, this type of error indicates a mistake of the annotation system itself and not a mistake in the modeling of the domain. As for the missing annotations, we can infer from the testing on annotation that it allows the lexical and conceptual enrichment of our ontology. Consequently, it could be possible to consider the annotation system as a tool to enrich ontologies.

Table 3 – Annotation results

Concepts...	Number
...annotated	724
...true positive (correct annotations)	607
...false positive (non-correct annotations)	117
...false negative (missing annotations)	134
Precision	83.8%
Recall	81.9%
F-mesure	82.8%

Although the experts were not easily available for validation, their involvement during the interviews was very useful and productive. Indeed, they were involved and interested in the project as well as in the task they were requested to perform. Conversely, when they were asked to proceed with the same validation assignment alone, without interaction with the ontologist nor with their colleagues, they did not show the same level of involvement.

Discussion

In our literature review about ontology validation, we found that all methods could be divided into methods to validate the structure and methods to validate the semantic. We aimed to define a method crossing these two challenges to have an end-to-end method. The most complex and time-consuming task of validation is the semantic validation by players of the domain, actors or experts. We found a solution using collaborative tools by organizing sessions of validation to involve them in the validation process. Eventually, the hardest challenge is less the interest of experts for the validation than the availability time of experts. We also spotlighted the importance of a testing phase in an application to validate the ontology adequacy, as well as to enrich the model.

The major observation from this study is the absolute necessity of domain players/experts' involvement. The domain players and experts are the owners of the academic and practical knowledge of their domain of expertise. For this reason, they cannot be replaced by automatic algorithm. Additionally, we recommend that, before planning ahead a modeling project, we make sure that the experts would be able to invest themselves in the project and allocate sufficient time to the related activities. The most challenging task we encountered was to involve the concerned experts, follow up on the time allocated to validate the ontology and increase their awareness regarding the time necessary to reach the quality they need in the targeted ontology. Therefore, establishing method based on interactions between the experts and the ontologist is essential and certainly save time and eases the process. After we set up the LOVMI methodology, we organized four validation sessions with six different players and took about three hours per player.

Conclusions

In this paper, we presented the development of a specific module related to social and environmental factors of mental disorders within ONTOPSYCHIA, the psychiatry ontology, and we mainly focused on the end-to-end validation methodology. First, we used a bottom up approach, extracting specialized terms from our corpus composed of Sainte Anne Hospital PDS. This first phase was concluded by the conceptualization of specialized terms into a domain ontology following the key aspects of METHONTOLOGY. During the second phase, we focused our work on the development of an end-to-end validation methodology. We established the LOVMI method,

providing a methodological framework to validate the structure and the semantics of any domain ontology.

This work confirmed the overarching aim of ONTOPSYCHIA, that, is to associate it with dedicated tools in order to: (1) perform semantic research in Patient Discharge Summaries (PDS), (2) represent the comorbidity, (3) index PDS for the constitution of cohorts and (4) identify resistant patients' profiles. We also postulate that this approach can help reach a consensus on descriptive categories of mental disorders.

We will add two modules to complete ONTOPSYCHIA: a module about mental disorders and a module about treatments. To develop the module about mental disorders, we chose a top down approach to ensure the alignment between current classifications. We aligned the DSM IV and the CIM 10 on the DSM V classification. We still have to validate this module following the LOVMI methodology. Regarding the development of module about treatments, we used a semi-automatic method, by extracting medication from an official list of marketing authorization. When these last modules will be built, we plan to assemble the three modules under the upper level ontology Basic Formal Ontology (BFO).

Acknowledgements

OntoPsychia is available following this address:
<https://bioportal.bioontology.org/ontologies/ONTOPSYCHIA>

References

- [1] J. Hastings, B. Smith, W. Ceusters, M. Jensen, K. Mulligan, Representing mental functioning: Ontologies for mental health and disease, *ICBO 2012: 3rd International Conference on Biomedical Ontology*, 2012.
- [2] Ceusters W, Smith B. Foundations for a realist ontology of mental disease. *J Biomed Semantics*. 2010 Dec 9;1(1):10. doi:10.1186/2041-1480-1-10. PubMed PMID:21143905; PubMed Central PMCID: PMC3017014.
- [3] J. Charlet, B. Bachimont, M.-C. Jaulent, Building medical ontologies by terminology extraction from texts: An experiment for the intensive care units. *Computers in biology and medicine* 36(7), (2006), 857–870.
- [4] S. Aubin, T. Hamon, Improving term extraction with terminological resources. *Advances in Natural Language Processing*, (2006), 380–387.
- [5] P. Denis, B. Sagot, Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. *Language resources and evaluation* 46(4), (2012), 721–736.
- [6] M. Fernández-López, A. Gómez-Pérez, N. Juristo, Methontology: from ontological art towards ontological engineering, *AAAI Symposium on Ontological Engineering*, Stanford,(1997), 33–40.
- [7] B. Bachimont, A. Isaac, R. Troncy, Semantic commitment for designing ontologies: A proposal, *Knowledge Engineering and Knowledge Management*, (2002), 114–121.
- [8] M. Poveda-Villalón, M.-C. Suárez-Figueroa, A. Gómez-Pérez, Validating ontologies with OOPS!, *Knowledge Engineering and Knowledge Management*, (2012), 267–281.
- [9] C. Grouin, A. Névoul, De-identification of clinical notes in french: towards a protocol for reference corpus development. *Journal of biomedical informatics* 50, (2014), 151–161.
- [10] J. Peace, P. Brennan, Ontological representation of family and family history. *AMIA Annual Symposium Proceedings*, (2007), 1072.
- [11] X. Aimé, J. Charlet, Preferred label validation by lexical prototypicality gradient: a use case on a rare diseases ontology, *Capturing and Refining Knowledge in the Medical Domain* (2012), 36.

Address for correspondence

Marion Richard
 LIMICS, Escalier D, 2eme étage
 15 rue de l'école de médecine, 75006 PARIS
marion.richard75@gmail.com

Surveying a New Multi-Institution Clinical Data Research Network

Marc Rosenman^{a,b}, Margaret Madden^a, Elissa Oh^a, Satyender Goel^{a,c}, Abel Kho^{a,c}

^a Center for Health Information Partnerships, Northwestern University, Chicago, Illinois, USA

^b Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

^c Department of Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

Abstract

Cultivated by the Patient-Centered Outcomes Research Network (PCORnet), thirteen regional clinical data research networks (CDRNs) are taking shape across the U.S. The PCORnet common data model was carefully planned, and the data marts assembled by the more than 80 data-contributing institutions (nodes) are undergoing, in 2016-2017, a series of data characterization cycles. PCORnet will adjudge each node's—and thereby, in a significant way, each CDRN's—readiness or unreadiness for multi-institution research. Certifying each node's quality and fidelity is of course essential. But in understanding network readiness there is an additional, vital dimension—one that has received too little attention. It is the development of knowledge about the nature of a CDRN's data, in its federated sense. With visualizations, how might one grasp the meta-data of a CDRN? We outline an approach that builds upon the HealthLNK Data Repository, a forerunner to the Chicago Area Patient-Centered Outcomes Research Network (CAPriCORN) CDRN.

Keywords:

Patient Outcome Assessment; Metadata; Electronic Health Records.

Introduction

Cultivated by the Patient-Centered Outcomes Research Network (PCORnet), thirteen regional clinical data research networks (CDRNs) are taking shape across the U.S [1-3].

PCORnet is part of the Patient Centered Outcomes Research Institute, which was authorized by the Patient Protection and Affordable Care Act of 2010. The PCORnet common data model was carefully planned, and the data marts assembled by the more than 80 data-contributing institutions (nodes) are undergoing, in 2016-2017, a series of data characterization cycles. PCORnet will adjudge each node's—and thereby, in a significant way, each CDRN's—readiness or unreadiness for multi-institution research. Certifying each node's quality and fidelity is of course essential. But in understanding network readiness there is an additional, vital dimension—one that has received too little attention. It is the development of knowledge about the nature of a CDRN's data, in its federated sense. With visualizations, how might one grasp the meta-data of a CDRN? We outline an approach that builds upon the HealthLNK Data Repository [4-6], a forerunner to the Chicago Area Patient-Centered Outcomes Research Network (CAPriCORN) CDRN [7,8].

The HealthLNK Data Repository is a de-identified assembly of electronic health records (EHR) of adults 18-89 years old, from seven Chicago health care institutions: five large academic medical centers, one large county health care system, and a network of community health centers. HealthLNK also created a software application to merge and de-duplicate the patient identifiers across the institutions [4]. Data from the seven

institutions are thereby woven together at the individual level via de-identified hashing. While the repository is maintained in a centralized system, housed in an enterprise data warehouse behind a secure firewall at Northwestern University, HealthLNK is a shared resource, to provide insight on the health of the Chicago community, and to identify opportunities to improve care. As a new repository it has proved efficient in providing data for studies of patterns of care among patients with these diverse conditions: diabetic ketoacidosis [6], diabetic retinopathy [9], gastrointestinal endoscopic procedures [10], systemic lupus erythematosus [12], and non-emergent conditions in the emergency department [12].

HealthLNK was designed by many of the same Chicago institutions that subsequently developed CAPriCORN, which is one of the thirteen PCORnet CDRNs. CAPriCORN includes the seven HealthLNK institutions plus four more. The hashing-and-matching software that HealthLNK developed for merging, de-duplicating, and de-identifying patient identifiers across the institutions has been adopted by CAPriCORN and some other CDRNs in PCORnet. For all of these reasons HealthLNK provides an opportunity to describe a CDRN at a federated level. How does one go about exploring a federated repository's strengths and weaknesses, and building its meta-data? The purpose of this activity is not to test a hypothesis but rather is to build pre-research knowledge. It may help investigators shape hypotheses and methods. It also may guide decisions about future sponsored opportunities to pursue.

Here we provide an overview of HealthLNK (as reflective of a nascent CDRN) from various angles:

- Contributions and interdigitations by data type:
 - Demographics
 - Diagnoses
 - Procedures
 - Laboratory results
 - Vital signs
 - Medications
- Overlap between the institutions' populations
- Emergency, inpatient, and outpatient encounters
- Subpopulations based on particular health conditions, or number of institutions visited
- Mortality
- Geographic scope

Methods

Retrospective data from 2006 to 2012 were available in the HealthLNK data repository. We used SQL queries to intersect the six principal data tables: demographics, diagnoses, procedures, laboratory results, vital signs, and medications.

For this Venn diagram we report the number of patients in the six-way and five-way intersections, and the single data type zones. We also determined in which institution(s) each patient had at least one diagnosis record. These diagnosis data are the substrate for a six-set Venn diagram, and some smaller diagrams focused on inpatient or emergency department encounters in a set of four institutions. In the analyses of diagnosis records we excluded one institution because it had data from only two years, and because a six-way Venn diagram is easier than a seven-way version to interpret visually. The template for our six-way Venn diagram is the one authored by Jeremy Carroll, PhD (then of Hewlett-Packard) [13].

Because CDRNs may have particular value in examining those patients who sought health care in more than one institution, we described additional dimensions of some scenarios for multi (or single) institution use, per patient. We examined institutional cross-over for patients with cancer, and for those who were the victim of a stabbing or gunshot wound. We also examined, among patients who had one or more diagnosis records in two institutions, how many diagnosis records the patient had from each institution. We also calculated the mortality rate by number of institutions visited.

The geographic analysis of Cook County (Chicago and nearby suburbs) and DuPage County (additional western suburbs), Illinois reflects the density of the HealthLNK population: number of unique patients in HealthLNK in each zip code divided by the total zip code population (U.S. Census). The map was drawn using ArcGis software (Esri, Redlands, CA). We included all seven institutions in the geographic analysis.

The geographic analysis of Cook County (Chicago and nearby suburbs) and DuPage County (additional western suburbs), Illinois reflects the density of the HealthLNK population: number of unique patients in HealthLNK in each zip code divided by the total zip code population (U.S. Census). The map was drawn using ArcGis software (Esri, Redlands, CA). We included all seven institutions in the geographic analysis. Encounter dates in HealthLNK are provided by the data-contributing institutions as MM/YYYY. HealthLNK's rules forbid comparison of the institutions by name.

Results

There are 3,697,707 unique patients in the 2006-2012 instance of the HealthLNK database (seven institutions). The numbers of unique patients by data type are shown in Table 1.

Table 1 – Patient Counts by HealthLNK Table

Data Table	Unique Patients with Data Type	% Database Population Total
Demographics	3,085,215	83%
Diagnosis	2,602,509	70%
Procedures	1,882,573	51%
Laboratory results	1,587,832	43%
Vital signs	1,955,212	53%
Medications	1,444,084	39%

Based on records in the Diagnosis Table (using the six institutions with more than two years of data), the distribution of number of distinct calendar months (with an encounter) per patient is shown in Figure 1. Across the 7-year retrospective period, about half of the patients had encounters in no more than two distinct calendar months. Patients by combination of data types are depicted in Figure 2. About one-sixth of the patients have all six data types, and slightly more than one-third have five of six data types. About one-fourth have only one data type (mostly demographics or vital signs); these single-type scenarios suggested

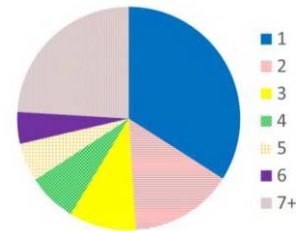


Figure 1 – Number of unique calendar months in which each patient had encounters (based on the Diagnosis Table)

that some data were omitted by one or two institutions in the extract/transfer/load procedures; their code will be revised in the next data refresh. Figure 3 shows the number of institutions visited per patient.

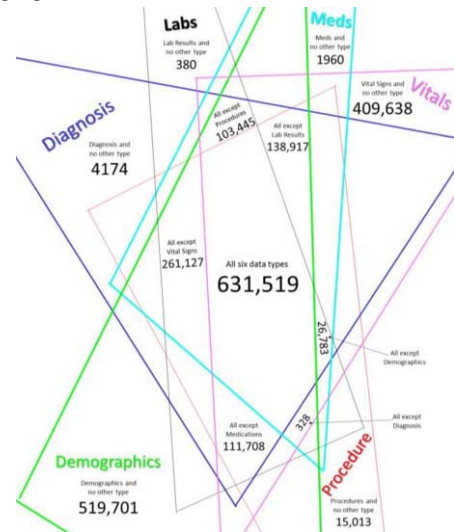


Figure 2 – Venn diagram (with truncated edges), six data types: the one, five, and six-way intersections are shown

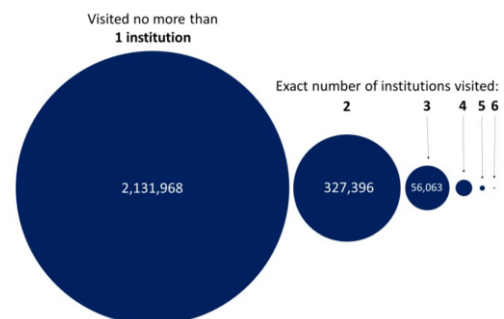


Figure 3 – Number of institutions visited per patient (based on the Diagnosis Table). Circles scaled by number of patients

Characteristics of patients by number of institutions visited are shown in Table 2. In the Diagnosis Table, we studied International Classification of Diagnosis, 9th Revision (ICD-9) codes, by number of institutions per patient. Among the small population with an encounter at all six institutions, more than 50% had an ICD-9 code for depression, and more than 40% for lack of housing. The frequencies of ICD-9 codes for drug abuse, chronic alcoholism, and suicidal ideation were also above 40%.

Table 2 – Patient Characteristics by N of Institutions Visited

Characteristic	Number of Institutions Visited				
	1	2	3	4	5
% Women	54	61	65	65	57
Race					
% Black	26	46	57	71	77
% White	49	41	30	22	20
% Other/declined	25	13	13	7	3
Insurance Status					
% Medicare	18	20	24	28	32
% Medicaid	7	13	15	20	25
% Commercial	57	39	29	21	13
% Self-Pay	11	19	21	21	22
% No Charge	0	0	1	1	1
% Other	8	9	10	9	7
Mortality (deaths/10,000)	1.7	5.4	4.8	7.7	12.6

* The denominator for this column is <100 patients.

When we examined the encounter types by “E,” “I,” and “O” (emergency department, inpatient, and outpatient, respectively), we noticed that one institution, which we know has a busy emergency department, had sent no type “E” but a very large number of type “I” records. Because of this apparent data transfer error, we excluded that institution from the “EIO” analyses. We also excluded the network of community health centers (all outpatient) and the institution with two years of data. Among the other four institutions, patient characteristics by number of institutions visited is shown in Table 3.

Table 3 – Patient Characteristics by N of Institutions Visited (Inpatient and Emergency Department)

Characteristic	Number of Institutions Visited				
Inpatient Institutions	1	2	3	4	5
% Women	49	62	59	46	26
Race					
% Black	50	26	40	51	53
% White	17	61	55	45	48
% Other/declined	33	13	5	4	0
Age (Median)	45	51	56	53	44
ER Institutions	1	2	3	4	5
% Women	61	57	64	59	41
Race					
% Black	50	43	70	76	67
% White	18	45	25	21	32
% Other/declined	32	12	5	3	1
Age (Median)	40	40	38	40	42.5

Characteristic	Number of Institutions				
Inpatient Institutions	0	1	2	3	4
% Women	49	62	59	46	26
Race					
% Black	50	26	40	51	53
% White	17	61	55	45	48
% Other/declined	33	13	5	4	0
Age (Median)	45	51	56	53	44
ER Institutions	0	1	2	3	4
% Women	61	57	64	59	41
Race					
% Black	50	43	70	76	67
% White	18	45	25	21	32
% Other/declined	32	12	5	3	1
Age (Median)	40	40	38	40	42.5

We also examined use of multiple institutions, within three subpopulations (again using the six institutions analyzed for Figure 3). A cohort with any malignancy diagnosis (which may include some patients who were evaluated and were found not to have cancer) was defined based on the presence of any ICD-9 diagnosis code in the range 140 to 209.99. A cohort with melanoma was defined based on at least one ICD-9 diagnosis code containing 172. A cohort with melanoma was defined based on at least one ICD-9 diagnosis code containing 172. A cohort with knife or gun-shot injury was defined by this set of ICD-9 E-codes: E965, E965.0, E965.1, E965.3, E965.4, E965.5, E965.8, E965.9, E966, E970, or E974. This analysis (Figure 4) considered visits of any type, for any diagnosis, among these three cohorts.

Those with any malignancy diagnosis are counted on the primary y-axis; those with melanoma or knife or gunshot injury on the secondary y-axis. Use of more than one institution was least among those with melanoma. Use of more than one institution was more common among those with knife or gun-shot injury.

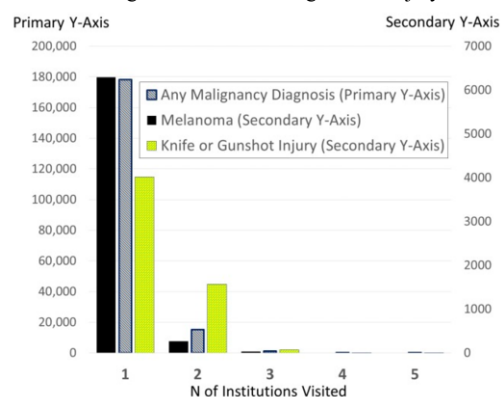
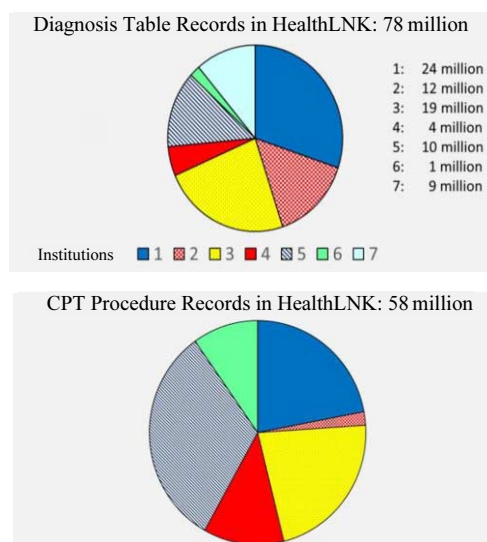


Figure 4 – N of institutions visited, among three cohorts

Apart from the issue noted above wherein one institution sent its type “E” encounters with a type “I” label, the Diagnosis Table and Procedure Table records sent to HealthLNK by the participating institutions were strong (Figures 5a and 5b).



Figures 5 – Distribution of Diagnosis (top) and Procedure (bottom) Table records, respectively, by institution

By contrast, not all of the institutions sent a full set of laboratory results. It can be helpful when visualizing a new CDRN to compare the distribution of institutions across different laboratory results. Figure 6 provides examples.

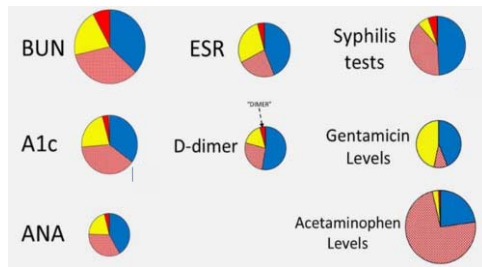


Figure 6 – Eight laboratory results by institution (BUN: Blood Urea Nitrogen; ESR: Erythrocyte Sedimentation Rate; A1c: Hemoglobin A1c; ANA: Anti-nuclear antibody)

In this way we discerned more than the fact that one institution provided only diabetes-related results. The distinctive institutional distributions of gentamicin levels and of acetaminophen levels reflects, it turns out, differences in actual care patterns (one of the hospitals in HealthLNK truly predominates in the number of patients evaluated with suspicion of acetaminophen toxicity) as well as technical (informatics) idiosyncrasies on the part of the data-supplying institutions.

The geographic distribution of the population (including all 7 institutions) in HealthLNK is illustrated in Figure 7. When we examined the patients with visits to two institutions, we found that in many cases, there existed only one (or a few) diagnosis record(s) in one of the two institution that the patient visited (Figure 8).

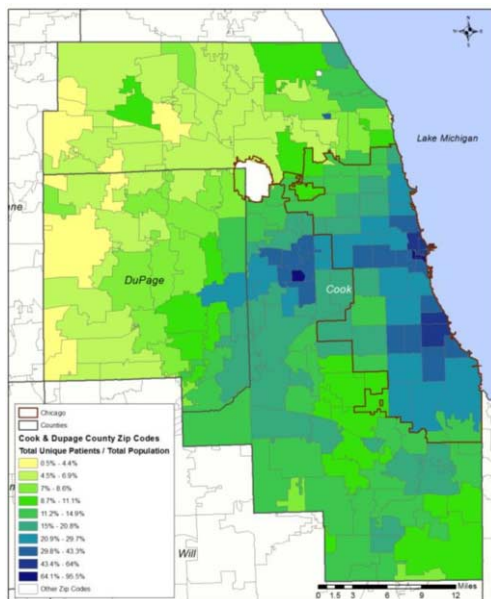


Figure 7 – Zip codes in darker blue have a larger population in HealthLNK (as a proportion of U.S. Census population)

More than 40% of these patients had only one or two diagnosis records in the second institution. About 33% of these patients had more than five diagnosis records in the second institution.

Discussion

In this survey of a forerunner to a PCORnet CDRN, we illustrated several dimensions in which it may be helpful to build understanding of the meta-data. In analyses after review of Table 1, we found that not all institutions sent laboratory data. One institution sent only diabetes-related laboratory results, in preparation for a specific research project. We then further learned about nuances by laboratory test. We also conducted indepth visualizations of the quality of various data elements in several domains (data not shown), such as variation between (and within) institutions in microbiology culture laboratory results.

Not unexpectedly, Figure 1 shows that most patients in a multi-institution, outpatient and inpatient EHR database have relatively few health care encounters over the years. Venn diagrams like that in Figure 2 are useful in planning potential studies of quality-of-care metrics. The zone with all (five) other data types, except for medication data, will be important to consider when calculating metrics in which the numerator is based on the use of a particular class of medication. Not unexpectedly, Figure 3 shows that the vast majority of patients visited no more than one institution over the years. Under 3% visited more than two institutions.

Table 2 shows that those who visited a higher number of institutions visited were more likely to men, African-American, and patients with publicly-financed or self-pay health care coverage. Mortality was strongly associated with number of institutions visited between 1 and 5. There were no deaths in the small group of patients who visited 6 institutions. The latter group had a very high rate of psychiatric diagnoses, substance abuse, and homelessness. Table 3 shows that among patients with at least one hospitalization during the study period, those who were hospitalized in multiple institutions over time were more likely to be men and to be African-American. The direction of these associations was the same for the number of emergency department institutions visited over the years.

Figure 4 suggests that the degree of “crossover” between institutions will have to be examined project by project, as patterns may well vary among different study cohorts. The contrasts between Figures 5a/5b and 6 suggest that additional exploration of (federated) meta-data for laboratory tests is needed.

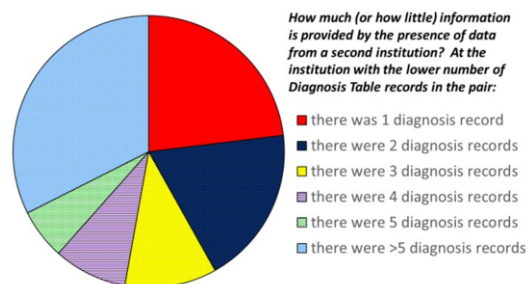


Figure 8 – Diagnosis records per patient in the one of the two institutions, among patients who visited two institutions

The geographic distribution in Figure 7 is consistent, in general, with the locations of the HealthLNK institutions. Within the city limits, HealthLNK’s population penetrance is lowest on the northwest and southwest sides of Chicago. Those neighborhoods are served, to a substantial degree, by hospitals that are not part of HealthLNK. In this context, it is worth considering what percentage HealthLNK is, of Chicago’s overall health care plant. Based on publically available data, we calculate that the HealthLNK hospitals account for approximately 40% of the acute care adult inpatient beds in the city, and approximately 30% of the labor and delivery beds. Figure 8 suggests additional health services research

work we may pursue in examining patients' use of more than one institution in a CDRN.

CDRNs and other multi-institution networks vary in geographic scope, types of constituent institutions, degree of centralization of data (and, hence, meta-data), etc. Those in a city (or any locale) might ask what proportion of total hospital beds in the area are within the network. The techniques we describe in examining overlaps of data types and populations may assist other networks in understanding care patterns and the network's relationship with its milieu. One might ask questions like: For newborns in a CDRN, to what extent does the CDRN have data on well-child visits in the ensuing years? What are the implications for child health research projects?

Networks can also build meta-data knowledge about particular data elements. In Table 2 above, we use the six categories of health insurance stored in HealthLNK. But what is the consistency and fidelity with which the institutions mapped their granular health insurance values to the six categories? We have begun comparing these mapping details and will build our meta-data knowledge to the point where we plan to offer new guidance to all networks in categorizing insurance.

The principal limitation of this project is that it is pre-research visualization rather than hypothesis driven work. Nevertheless, we think that this type of work is necessary—and needs more attention paid to it—in order to support the more specific research projects that will follow.

Conclusion

The HealthLNK project has substantially informed the work now underway in the CAPriCORN CDRN. In turn, the work in both CAPriCORN and PCORnet will help standardize and will enhance future versions of the HealthLNK repository.

Acknowledgements

This study was not sponsored.

References

- [1] D.A. Corley, H.S. Feigelson, T.A. Lieu, E.A. McGlynn, Building data infrastructure to evaluate and improve quality: PCORnet, *J Oncol Pract* **11** (2015):204-206.
- [2] L.H. Curtis, J. Brown, R. Platt, Four health data networks illustrate the potential for a shared national multipurpose big-data network, *Health Aff (Millwood)* **33** (2014): 1178- 1186.
- [3] C. Parke, J. Cook, T. Carton, S. Rao, The Louisiana clinical data research network: leveraging regional and national resources to improve clinical research efficiency, *Ochsner J* **14** (2014):718-723.
- [4] A.N. Kho, J.P. Cashy, K.L. Jackson, A.R. Pah, S. Goel, J. Boehnke, J.E. Humphries, S.D. Kominers, B.N. Hota, S.A. Sims, B.A. Malin, D.D. French, T.L. Walunas, D.O. Meltzer, E.O. Kaleba, R.C. Jones, W.L. Galanter, Implementation linkage tool in Chicago. *J Am Med Inform Assoc* **22** (2015):1072-1080.
- [5] W.L. Galanter, A. Applebaum, V. Boddipalli, A. Kho, M. Lin, D. Meltzer, A. Roberts, B. Trick, S.M. Walton, B.L. Lambert, Migration of patients between five urban teaching hospitals in Chicago, *J Med Syst* **37** (2013), 1-8.
- [6] J.A. Mays, K.L. Jackson, T.A. Derby, J.J. Behrens, S. Goel, M.E. Molitch, A.N. Kho, A. Wallia, An evaluation of recurrent diabetic ketoacidosis, fragmentation of care, and mortality across Chicago, Illinois, *Diabetes Care* **39** (2016):1671-1676.
- [7] A. Solomonides, S. Goel, D. Hynes, J.C. Silverstein, B. Hota, W. Trick, F. Angulo, R. Price, E. Sadhu, S. Zelisko, J. Fischer, B. Fumer, A. Hamilton, J. Phua, W. Brown, S.F. Hohmann, D. Meltzer, E. Tarlov, F.M. Weaver, H. Zhang, T. Concannon, A. Kho, Patient-centered outcomes research in practice: the CAPriCORN infrastructure, *Stud Health Technol Inform* (2015):584-588.
- [8] A.N. Kho, D.M. Hynes, S. Goel, A.E. Solomonides, R. Price, B. Hota, S.A. Sims, N. Bahroos, F. Angulo, W.E. Trick, E. Tarlov, F.D. Rachman, A. Hamilton, E.O. Kaleba, S. Badlani, S.L. Volchenboum, J.C. Silverstein, J.N. Tobin, M.A. Schwartz, D. Levine, J.B. Wong, R.H. Kennedy, J.A. Krishnan, D.O. Meltzer, J.M. Collins, T. Mazany, CAPriCORN Team, CAPriCORN: Chicago area patient-centered outcomes re- search network, *J Am Med Inform Assoc* **21** (2014):607-611.
- [9] D.D. French, J.J. Behrens, K.L. Jackson, A.N. Kho, T.L. Walunas, C.T. Evans, M. Mbagwu, C.E. Margo, P.J. Bryar, Payment reform needed to address health disparities of undiagnosed, *Ophthalmol Ther* (2016):123-131.
- [10] K.L. Jackson, S. Goel, A.N. Kho, R.N. Keswani, Distance from hospital impacts adverse event detection after outpatient endoscopy, *Gastrointest Endosc* **85** (2016):380-386.
- [11] T.L. Walunas, K.L. Jackson, A.H. Chung, K.A. Mancera-Cuevas, D.L. Erickson, R. Ramsey-Goldman, A. Kho, Disease outcomes and care fragmentation among patients with systemic lupus erythematosus. *Arthritis Care Res (Hoboken)* (2016) [Epub ahead of print]
- [12] J. Fishman, S. McLafferty, W. Galanter, Does spatial access to primary care affect emergency department utilization for nonemergent conditions? *Health Serv Res* (2016). [Epub ahead of print]
- [13] Diagrams Made from Triangles. [The Electronic Journal of Combinatorics website, edited June 2005] Accessed 12/23/2016. Available from: <http://www.combinator-ics.org/files/Surveys/ds5/VennTriangleEJC.html>.

Address for correspondence

Marc Rosenman, MD

625 N. Michigan Avenue, 15th Floor Chicago, IL, 60611, USA

Email: marc.rosenman@northwestern.edu

Phone: 312-503-3738

Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration

Stefan Schulz^a, Jean-Marie Rodrigues^{b,c}, Alan Rector^d, Christopher G. Chute^e

a IMI, Medical University of Graz, Austria, *b* INSERM U1142, LIMICS, Paris, France

c Department of Public Health and Medical Informatics, Univ. Jean Monnet of Saint Etienne, France

d University of Manchester, United Kingdom

e Division of General Internal Medicine, Johns Hopkins University, Baltimore, MD, 21287, USA

Abstract

The time has come to end unproductive competitions among different types of biomedical terminology artefacts. Tools and strategies to create the foundation of a seamless environment covering clinical jargon, clinical terminologies, and classifications are necessary. Whereas language processing relies on human interface terminologies, which represent clinical jargon, their link to reference terminologies such as SNOMED CT is essential to guarantee semantic interoperability. There is also a need for interoperation between reference and aggregation terminologies. Simple mappings between nodes are not enough, because the three kinds of terminology systems represent different things: reference terminologies focus on context-free descriptions of classes of entities of a domain; aggregation terminologies contain rules that enforce the principle of single hierarchies and disjoint classes; interface terminologies represent the language used in a domain. We propose a model that aims at providing a better flow of standardized information, addressing multiple use cases in health care including clinical research, epidemiology, care management, and reimbursement.

Keywords:

Terminology as Topic; Dictionaries as Topic; Knowledge Bases

Introduction

The evolution of electronic health records has been accompanied by the development of numerous and increasingly sophisticated lexical, terminological and ontological tools and resources, supporting structured data entry and processing of unstructured narratives [1]. A major challenge is to preserve meaning from descriptions of individual subjects of care to descriptions of populations, as well as from medical care to biomedical research. Attempts to do so have often neglected the fundamental reconciliation between different genres of biomedical terminologies used for these different purposes and use cases [2]. These genres include ontologies, reference terminologies, interface terminologies, and classifications (in this paper referred to by the more general term "aggregation terminologies" [3]). The reconciliation approach in [4] proposes to address this issue in two steps, links to and from:

- Interface terminologies and reference terminologies

- Reference terminologies and aggregation terminologies.

The two-step approach is required to prevent "confusion of concepts and the words used to express those concepts" [5], which is essential in order to achieve better interoperability at a time where clinicians, documentation specialists, epidemiologists, health care administrators, payers and health service researchers increasingly require that clinical data captured at one place be processed and analysed in different application contexts.

Methods

In the following we analyse the different terminology genres (interface, reference, and aggregation terminologies) and provide arguments that justify this distinction.

Interface vocabularies

Mainstream work on terminology and ontology during the last twenty years has been guided by a normative perspective, primarily driven by the English speaking community. Apart from increasingly incorporating principles of Applied Ontology [6] into terminologies and thesauri (with the Gene Ontology [7] and SNOMED CT [8] being the most prominent examples), the labelling of the nodes in these system has mainly followed a top-down strategy, with naming conventions emphasizing maximally self-explanatory and unambiguous labels such as "Malignant tumour of thyroid gland (disorder)". However, these labels, as clearly understandable as they are, do often not represent the language used by clinicians (e.g., "Thyroid Ca").

This gap is typically filled by (human) interface terminologies [9], i.e. collections of language expressions that actually occur in medical documentation. Such interface terms are typically the building blocks of clinical narratives but also are used as text values for structured data entry. However, there are several issues with interface terms, which often make them unsuitable for labelling reference terminology content:

- Interface terms tend to be as short as possible, and therefore ambiguous out of context. Abbreviations and acronyms play a major role, e.g. "CA" may mean "calcium", "cancer", and "cholic acid".
- Interface terms have different meanings in different user groups, characterised by medical professions and medical specialties, regional dialects and geo- graphic

names (e.g. "GWB": "general well-being", but in New York hospitals also: "George Washington Bridge").

- The meaning of interface terms may change across time, e.g., the acronym "AIDS" has been used, for a long time, for "Acquired Immunodeficiency Syndrome", although other expansions such as "Acquired Iatrogenic Death Syndrome" can be found in the literature.

As a consequence of the dynamics of clinical and scientific language, good interface terminologies require continuous maintenance. Interface terms need to be harvested from "living" language sources. They need to be set in a context, which makes their different meanings transparent, e.g. "Ureter Ca", "Ca level", instead of just "Ca". Only under these conditions, they can be reliably anchored within reference terminologies.

Reference terminologies

In contrast to interface terminologies, reference terminologies should provide stable and well-defined representational units (aka "concepts", "classes", "descriptors" or – confusingly – "terms"). The stability of these units relies not only on unambiguous textual labels, but also on textual definitions or scope notes, links to external standards, as well as on formal-ontological definitions usually based on Description Logics [10], typically using or referring to the OWL [11] language, like in SNOMED CT. E.g., the SNOMED CT concept Pancreatitis is defined as being logically equivalent to a disorder with inflammatory morphology that is located at some pancreas structure.

Connection of reference terminologies with other terminology system must address epistemically-"infested" content [13], i.e. reference to a concept within a discourse context that expresses negation, doubt, intention or risk. Interface terminologies may include terms like "suspected leukaemia" (in some languages like German even fused in a single term, "Leukämieverdacht"). The same is found in aggregation terminologies like ICD-10 "Glaucoma suspect" (H40.0) or "Observation for suspected tuberculosis" (Z03.0). This requires that the reference terminology provides a mechanism to deal with epistemic contexts. SNOMED CT's attempt to this is the context model (the "Situation with explicit context" hierarchy branch), which, however, exhibits several weaknesses under ontological scrutiny [14]. An ontologically founded model to represent both clinical entities and information entities and to connect them with each other was proposed in SemanticHealthNet [15]. This approach used OWL expressions to describe the compositional structure of information models, all of which under the BioTopLite class "information object". The relation between information objects and clinical entities or classes thereof is done via the object property "represents".

Aggregation terminologies

Aggregation terminologies contain rules that enforce the principle of single hierarchies and disjoint classes. This makes them mostly suited for statistical analyses. The most important aggregation terminology is the International Classification of Diseases.

Our experience of linking reference terminologies with aggregation terminologies is based on the ICD – SNOMED CT harmonization process, including a preliminary SNOMED CT – ICD 10 mapping based on expert knowledge of both coding and medicine [16], and on extensive work on the 11th revision

of ICD [17]. ICD-11 was designed on top of a multi-component architecture [18]. We here focus on the component to be released first, probably in 2018, viz. the Mortality and Morbidity Statistics [19].

Fig.1 shows the three building blocks of a terminology ecosystem constituted by clinical language resources (left), reference terminologies with or without ontological foundation (centre) and the building blocks of advanced or aggregation terminologies.

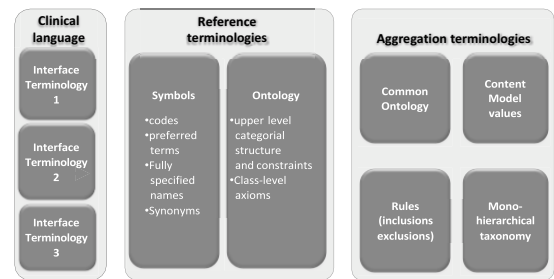


Figure 1- Three building blocks of a terminology ecosystem constituted by clinical language resources (left), reference terminologies with or without ontological foundation (centre) and building blocks of (advanced) aggregation terminologies.

Results

In this section we present the results of our scrutiny of current terminology systems, propose a general typology and provide some recommendations for their further evolution.

Lack of interface terms in reference terminologies

The need of interface terms becomes obvious when matching terminologies with clinical narratives such as clinical notes or discharge summaries. In an experimental study on the coverage of English and Swedish SNOMED CT releases used as annotation vocabulary for a corpus created out of hybrid clinical document samples in several languages [21], a nearly equivalent rate of concept coverage (87%) contrasted with a neatly different, and generally lower coverage of terms; with 47% for Swedish compared to 68% for English. This difference is due to the fact the Swedish SNOMED CT has only one term per concept, whereas the English SNOMED CT version has more than two, on average. These results demonstrate the need for language-specific interface terminologies, and they put in question whether a simple enrichment of a clinical reference terminology with interface terms, as practiced by IHTSDO for English and Spanish is really sufficient.

Local interface terminology efforts are needed

We support ongoing national terminology building efforts as decentralised bottom-up activities, starting with a systematic collection of commonly used words and phrases in daily communication between patients and health professionals. For instance, an effort to build an interface terminology for German linked to SNOMED CT codes semi-automatically has already resulted in more than 1.8 million interface terms, not including short forms like acronyms and abbreviations for which methods of disambiguation and resolution are currently tested [22]. If such efforts are costly, they are more helpful for guaranteeing the use of reference terminologies and the seam-

less flow of meaningful clinical information than huge top-down reference terminology translation efforts.

It can also be useful to relate interface terminologies with thesauri like MeSH. Thesauri lack formal-ontological foundation [12], but provide precise textual definitions. Here, Pancreatitis is described by the scope note "Inflammation of the pancreas". Although such a description does not use a formal language it may support a mapping to an ontology-based reference terminology such as SNOMED CT.

General typology of terminology genres and interfaces

Based on our experience we propose the following distinction between terminology systems to integrate the three terminology blocks and the two interfaces between the blocks:

(First order) axiom-based systems – generally using description logics (DLs), which provide axioms for sub-class / super-class relations and existential restrictions (e.g. "every instance of A is located in some instance of B"). From these axioms, class hierarchies may be inferred algorithmically by DL reasoners [23]. Typically, this leads to poly-hierarchies. The semantics are that all statements are necessarily true in all possible interpretations. Such systems are open-world [10], and negation means necessarily false, i.e. false in all possible interpretations. No exceptions are allowed. All statements are first order, i.e. about all individuals in a class; statements about the classes themselves are not allowed.

Closed world systems – e.g. logic programming and database systems – have in common that their semantics is based only on what is held explicitly in the system – hence closed world. If hierarchies are present, they must be stated explicitly and cannot be inferred. False means "not provable in the closed world of the system" Therefore, new information about that world can falsify previous conclusions; exceptions may therefore be allowed. Consequently, universal subclass-superclass relations that must hold in any world cannot be inferred, because closed world statements can be proved in the closed world, they just cannot be proved universally for any world. This is typical of rule-based systems. They prescribe what to do in particular situations e.g., in languages for decision support systems. They are the foundation for mono-hierarchical aggregation terminologies and for queries on representations: Mono-hierarchical aggregation terminologies (aka statistical classifications), characterized by single hierarchies and disjoint classes, supported by a large corpus of exclusion and inclusion rules, as well as by coding guidelines, which vary between use cases (e.g. coding for reimbursement vs. coding for mortality statistics) and local contexts. Examples are ICD-9 and ICD-10, the upcoming ICD-11 linearization(s), other WHO classifications, and national catalogues (like, e.g. the German OPS procedure classification) [24].

Queries on representations – which may be used to extract information about any of the above kinds of representation artefacts – as opposed to the knowledge represented. The query languages SPARQL [25] with its specialised extension for DLs and the SNOMED CT Expression Constraint Language [26] are particularly important in medical applications for linking axiomatic systems to aggregation terminologies, because they support formalizations of the characteristic residual classes not classified under and not elsewhere classified. Examples for exclusions are the ICD-10 classes under I10-I15, characterised as "Hypertensive diseases, excluding complicating pregnancy, childbirth and the puerperium".

Knowledge Organization Systems (or "terminologies" in general) provide hierarchies based on loosely defined "broader-

than" / "narrower-than" relations between terms or groupings thereof, which are first of all seen as streamlining the navigation among human language terms for which also synonymy and other semantic relations can be asserted. The most popular knowledge organization system in the field of biomedicine is the MeSH thesaurus [12] with its multiple tree structure and entry terms. In MeSH, negation has no formal meaning, nor is it built upon any explicit ontological foundation.

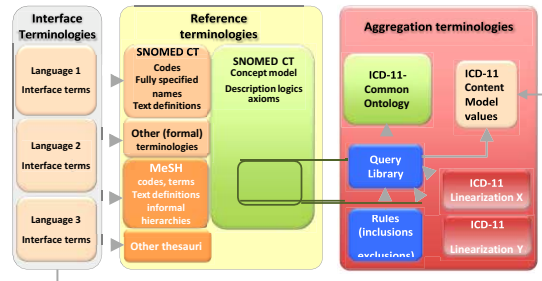


Figure 2- Architecture from Fig. 1, applied to SNOMED CT and ICD-11, with one or more linearizations (e.g. the ICD-11 Mortality and Morbidity Linearization [19] as example of an aggregation terminology as a possible end product. The ICD-11 common ontology is a subset of SNOMED CT. Linearizations are built using language-specific labels from interface terminologies (value sets), which can also be extended by clinical language synonyms, e.g. by additional thesauri like [20]. The colours have the following meaning: Green: Axiom based systems; Pink: Closed world systems including dictionaries; Orange: knowledge organization systems; Blue: queries on representations and annotations; Red: Mono-hierarchical aggregation terminologies .

Fig. 2 summarizes the links between three blocks of a specific terminology ecosystem, with different kinds of resources and technologies, regarding the content of SNOMED CT, a subset of it that qualifies as common ontology for ICD-11, and additional resources that secure the architectural principles of aggregation terminology linearizations.

In closed-world systems like ICD-x and other aggregation terminologies it can be asserted that something is true in the world of the representation. For instance, it is sufficient in the above example, to assert a hypertensive disease code as long as there is no evidence that the patient is pregnant or in the perinatal phase. That is, they are usually true, or true under certain conditions, but not necessarily true by definition. The latter would occur when using axiom-based systems were used for axiomatising the content of aggregation terminologies: Using logical negation for representing hypertensive disease as above would entail that any patient classified as hypertensive was not pregnant or in the perinatal phase. This kind of entailment is not intended by aggregation terminologies.

Therefore, if axiom-based systems such as SNOMED CT are to be linked to aggregation terminologies such as ICD-x, queries on representations are needed. It is the underlying axiom-based system in description logic that allows SNOMED CT to fulfil the twelve Cimino criteria for controlled vocabularies [27, 28]. By its Expression Constraint Language, SNOMED CT provides a means of formulating queries on that representation to bridge between its reference terminology and aggregation terminologies.

In the case of the link between SNOMED CT and ICD-11 Mortality and Morbidity Statistics version [19] as one of different possible end products, the ICD-11 Common Ontology

(Fig.2) is a subset of SNOMED CT. Linearizations are built using language-specific labels from specific interface terminologies (value sets), which can also be extended by clinical language synonyms, e.g. by additional thesauri like [20].

For interface terminologies there is currently no agreement on which formalism to use to anchor them on reference terminologies. A possible Semantic Web standard is SKOS [29], currently aligned with ISO standards for thesauri. Interface terminologies supporting aggregation terminologies, but also value sets that provide canonical names raise another issue, viz. imprecision. Interpreting these terms (like “Diabetes mellitus”) at face value, is not consistent with their exact meaning, such as Diabetes mellitus in patients that are not pregnant or in the perinatal phase. Shortcut concept “alignments” that firstly rely on lexical criteria and which (falsely) infer equivalence of meaning from string equivalence of identifiers, are therefore inherently imprecise.

Conclusions

A major difficulty in reconciling interface terminologies, reference terminologies (e.g., SNOMED CT) and aggregation terminologies (e.g., ICD-11) was understanding the roles of the different terminology types and how they are related to the distinctions between (i) kinds of terminological/ontological knowledge, (ii) meaning of language expressions, (iii) the things they denote, (iv) necessary truths about classes of things, and (v) use case specific interface terms, explanations and rules. Our suggested typology of representational artefacts could help prevent difficulties in specifying terminology architectures like the one underlying ICD-11. In order to ease understanding of the distinctions between statements of necessary truths, closed world knowledge, navigation associations, and classifications rules, we suggest the following vocabulary for a component-based architecture as sketched in Fig.2:

- The open world component comprising first-order necessary truths, thus constituting the ontological basis of reference and aggregation terminologies in the near future.
- The closed-world component includes rules that assure the architectural constraints of aggregation terminologies. It is the foundation of two other components:
 - Aggregation terminologies, i.e., classification systems in a broad sense, are constituted by single hierarchies for specific purposes, like statistical reporting and billing. They follow the jointly- exhaustive-mutually-exclusive rule.
 - Query libraries express the meaning of nodes of aggregation terminologies, by querying against reference terminologies.
- The knowledge organization foundations include all other kinds of supportive, more loosely specified knowledge resources including lexicons with interface terms. It provides the framework for multiple interface terminologies to be implemented as dynamic plugins for different languages and communities.

Acknowledgements

This work was supported by the World Health Organization (WHO) and SNOMED International (IHSTDO) through their Joint Advisory Group (JAG). It also refers to work that received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 643818.

References

- [1] O. Bodenreider. Lexical, terminological and ontological resources for biological text mining. S. Ananidou et al, Text mining for biology and biomedicine; Artech House, London, UK, 2006, 43-66.
- [2] C. G. Chute. Clinical classification and terminology. *JAMIA* 7;3 (2000), 298-303.
- [3] J. Rogers. Using Medical Terminologies. 2005, <http://www.cs.man.ac.uk/~jeremy/HealthInf/RCSEd/terminology-using.htm>
- [4] R. H. Baud, W. Ceusters, P. Ruch, A. M. Rassinoux, C. Lovis, A. Geissbühler. Reconciliation of Ontology and Terminology to cope with Linguistics, *Studies in Health Technology and Informatics* 129 (2007), 796-801.
- [5] A. L. Rector, Clinical Terminology: Why is it so Hard? *Methods of Information in Medicine*, 38, (1999), 147-157.
- [6] B. Smith. Applied Ontology. A new discipline is born. *Philosophy Today* 12,29 (1998), 5-6
- [7] M. Ashburner et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nature Genetics* 25;1 (2000),25-9.
- [8] SNOMED International (aka International Health Terminology Standards Development Organisation - IHSTDO). SNOMED CT – The Global Language of Healthcare (2016), <http://ihtsdo.org/snomed-ct>
- [9] S.T. Rosenbloom, R.A. Miller, K.B. Johnson, P.L. Elkin, S.H. Brown. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *JAMIA*13;3 (2006), 277-288
- [10] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. *The description logic handbook*, second edition 2, Cambridge University Press, Cambridge, UK, 2010
- [11] World Wide Web Consortium (W3C). OWL2 Web Ontology Language (2012). <https://www.w3.org/TR/owl2-overview/>
- [12] U.S. National Library of Medicine. Medical Subject Headings (MeSH) (2016) <https://www.nlm.nih.gov/mesh/>
- [13] O. Bodenreider, B. Smith, A. Burgun: The Ontology- Epistemology Divide: A Case Study in Medical Terminology. 3rd Conference on Formal Ontology in Information Systems (2004), 185–195.
- [14] C. Martínez-Costa, S. Schulz. Ontology-based reinterpretation of the SNOMED CT context model. *International Conference on Biomedical Ontologies (ICBO)* (2013) 90-95.
- [15] C. Martínez-Costa, R. Cornet, D. Karlsson, S. Schulz, D. Kalra. Semantic enrichment of clinical models towards semantic interoperability. The heart failure summary use case. *JAMIA* 22;3 (2015), 565-576.
- [16] K. Giannangelo, J. Millar, Mapping SNOMED CT to ICD-10, *Studies in Health Technology and Informatics* 180, (2012), 83-87.
- [17] The World Health Organization, The International Classification of Diseases 11th Revision is due by 2017, <http://www.who.int/classifications/icd/revision/en/>
- [18] J.M. Rodrigues, D. Robinson, V. Della Mea, J. Campbell, Rector, S. Schulz, H. Brear, B. Üstün, K. Spackman, C., G. Chute, J. Millar, H. Solbrig, K. Brand Persson. *Studies in Health Technology and Informatics*, 216 (2016), 790- 794.
- [19] M. Mamou, A. Rector, S. Schulz, J. Campbell, H. Solbrig, J.M. Rodrigues. Representing ICD-11 JLMMS Using IHSTDO Representation Formalisms. *Studies in Health Technology and Informatics* 228 (2016), 431-435.
- [20] DIMDI (German Institute for Medical Documentation and Information). ICD-10 Thesaurus of diagnostic terms (IDT). <https://www.dimdi.de/static/en/klassi/icd-10-gm/historie/idt.htm>
- [21] ASSESS CT. Assessing SNOMED CT for Large Scale eHealth Deployment in the EU. <http://assess-ct.eu>
- [22] S. Schulz, J.A. Miñarro-Giménez. Using Language Technology for SNOMED CT Localisation. *IHSTDO Showcase* (2015), Montevideo, Uruguay
- [23] N. Matentzoglou, J. Leo, V. Hudhra, U. Sattler, B. Parsia. A survey of current, stand-alone owl reasoners. *Informal Proceedings of the 4th International Workshop on OWL Reasoner Evaluation*, Vol 1387 (2015)
- [24] DIMDI (German Institute for Medical Documentation and Information). OPS – German Procedure Classification. <https://www.dimdi.de/static/en/klassi/ops/index.htm>

- [25] W3C Consortium. SPARQL Query Language for RDF(2008)
<https://www.w3.org/TR/rdf-sparql-query/>
- [26] IHTSDO. SNOMED Expression Constraint Specification and Guide (Draft). 2016.
- [27] J.J. Cimino, From Data to Knowledge through Concept- oriented Terminologies: Experience with the Medical Entities Dictionary, JAMIA 7;3 (2000), 288-297.
- [28] J.J. Cimino, In defense of the desiderata, Journal of Biomedical Informatics 39;3 (2006), 299-306.
- [29] W3C. SKOS – Simple Knowledge Organization System (2012),
<https://www.w3.org/2004/02/skos/>

Address for correspondence

Stefan Schulz

Institute for Medical Informatics, Statistics and Documentation

Medical University of Graz

Auenbruggerplatz 2/V A- 8036 Graz (Austria)

stefan.schulz@medunigraz.at

Semantic Web Service Delivery in Healthcare Based on Functional and Non-Functional Properties

Marco Schweitzer, Thilo Gorfer, Alexander Hörbst

eHealth Research and Innovation Unit, UMIT - University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

Abstract

In the past decades, a lot of endeavor has been made on the trans-institutional exchange of healthcare data through electronic health records (EHR) in order to obtain a lifelong, shared accessible health record of a patient. Besides basic information exchange, there is a growing need for Information and Communication Technology (ICT) to support the use of the collected health data in an individual, case-specific workflow-based manner. This paper presents the results on how workflows can be used to process data from electronic health records, following a semantic web service approach that enables automatic discovery, composition and invocation of suitable web services. Based on this solution, the user (physician) can define its needs from a domain-specific perspective, whereas the ICT-system fulfills those needs with modular web services. By involving also non-functional properties for the service selection, this approach is even more suitable for the dynamic medical domain.

Keywords:

Electronic Health Records; Semantics; Technology

Introduction

With the advent of electronic health records (EHRs) the shared trans-institutional data exchange in healthcare has led to an unprecedented amount of health data in the past decade. Today facilitating this data in a meaningful way is crucial for high quality healthcare delivery [1].

Related information and communication technology (ICT) solutions are necessary to manage this emerging data and use it for purposeful and individual diagnosis and therapy. In general, medical diagnosis and treatment follow a certain workflow, i.e. a sequence of tasks (both, IT-based and human-based) is executed by an individual or a group of healthcare professionals to obtain some business goal. With the means of business process modeling, such a workflow can be modeled graphically using business process diagrams (BPD), e.g. with the business process model and notation (BPMN). The latest version 2.0 of BPMN extended the solely graphical tool with capabilities to execute process diagrams through workflow engines. This opens new ways of integrating BPMN 2.0 with ICT, in order to improve and automate business processes. Up to 50% of everyday routine tasks for diabetes consultations belong to recurring IT-related tasks [2], e.g. finding reports, change a therapy plan or organize a referral. However, there is little experience in automating such clinical processes. These tasks could be modeled as BPD at point of care by the user (e.g. a physician) and used for automatic execution to support routine work.

Today's EHR systems have two major limitations:

1. They solely allow access to raw data. Standards like HL7 CDA [3] are very comprehensive and new approaches like HL7 FHIR [4] allow dynamic and flexible data access but the main focus is set to raw data retrieval.
2. ICT Systems in healthcare tend to be designed in a monolithic manner, i.e. a well-designed and conceptualized problem-scope is solved by using defined, static functions with limited extensibility. Changes within the software design require reengineering the software, which is correlated to increased effort and costs. Further no individual customization of information processing (data processing functions) is permitted, though present publications highlight challenges and burden relating EHR for routine workflows [5,6].

So the design and implementation of meaningful healthcare software systems need to offer comprehensive functionality based on individual preferences and customized EHR workflows and data flows [7] on one side, and on the other side support adaption to the highly dynamic and quickly evolving medical domain by providing scalable, distributed and interoperable systems [8].

Existing research on EHR related workflows have mostly focused on clinical pathways and outline the urgent need for EHRs to comply with workflows [9]. The project *OntoHealth*¹ aims at developing concepts and systems, that are able to process EHR-related workflows based on BPMN 2.0, which are executed using semantic web services. Very important and often not sufficiently considered criteria for present software systems are given through non-functional properties (NFPs). There are approaches that enhance BPMN with NFPs in order to establish quality conditions on tasks [10], but according to the knowledge of the authors, no literature could be found that try to extend EHR data processing for user defined workflows with semantic web services incorporating NFPs. In the *OntoHealth* project NFPs play a crucial role as they define the desired quality parameters for service selection and are used for proper service execution. Thus, besides deciding what data from an EHR is needed and how it should be processed by suitable web services, the user can select NFPs (e.g. performance, costs), which define quality constraints and certain service selection needs to adhere when executing a workflow.

This paper describes our approach of how we automatically align user defined EHR workflows represented as a sequence of tasks with semantic web services under consideration of NFPs.

¹ <http://www.ontohealth.org>

Methods

Today prominent semantic standards like the resource description framework (RDF), the web ontology language (OWL) and SPARQL provide suitable means to use semantic technologies even in the medical domain, e.g. to structure medical knowledge like SNOMED-CT. For EHRs, semantic technologies are crucial to achieve semantic interoperability. However, the use of semantics for clinical functionality is limited.

A key technology in today's software development is the modular approach of utilizing loosely coupled web services. In so called Service Oriented Architectures (SOAs) several modular functions of a software are outsourced from its main application and processed on isolated services. Those loosely coupled and high cohesive modules are aligned to fulfill certain business needs. The web services are usually realized with standards like SOAP or the very lightweight REST standard.

Beside the way of directly calling a known service (e.g. calling the service RiskCalculator provided by hospital A with parameters XYZ) it is possible to find a certain service from a service repository fulfilling one's business needs. Usually SOAs provide some kind of service broker that is responsible to register several services from different providers in such registries, where each service includes a description document or service contract (JSON or XML) that contains meta information about the preconditions, capabilities (inputs/outputs, effects) and interfaces of the service. The service broker also answers requests of the service consumers, e.g. with the information on how to call the interface of the service. The limitation here is that common web service description documents don't share any semantic description and thus are not machine readable. For this purpose, Semantic Web Services (SWS) provide the means to combine semantic technologies with web services: Each service owns a description that is enriched with semantic content, in order to automate service discovery, composition, invocation and monitoring.

In our approach we adapted the service delivery structure of WSMO (Web Service Modelling Ontology) [11]. Based on this concept we integrated our developed domain ontologies for EHR web services, containing pertinent data elements as in- and outputs, functions and NFPs for EHR data processing. The service contract is modeled using MSM (Minimal Service Model) with a slight extension for the integration of data elements and functions. Our idea was to use a direct² matching based alignment of functions, inputs, outputs and NFPs from user defined tasks and available web services. We build a service delivery pipeline which covers the following web service delivery steps: discovery, composition, ranking and invocation. For our purpose we aligned those steps with the domain ontology needs and the workflow-based execution, i.e. the steps deal with functions, inputs, outputs and NFPs in a suitable way. Further parts for registration and pre-filtering were added to provide complete service delivery. Rules were defined on how to enable interaction between the steps of the pipeline process and on how to match the related requirements (functions, inputs, outputs, NFPs).

The developed service delivery process was implemented as a prototype application in order to prove the concept.

Results

Within the OntoHealth project, a workflow-based SOA platform has been conceptualized, which executes an individual,

problem-based process (modeled with BPMN 2.0) by means of an automatic semantic web services composition. The overall architecture is presented [12]. In this paper, we present the precise execution flow with details on how the different modules are combined to get the related services. An overview of the steps is depicted in figure 1. A given workflow consists of known flow objects (events, activities and gateways). As defined in BPMN a task is a sub element of an activity and for our purpose contains the information about the user defined requirements. Those requirements comprise for each task inputs and outputs (data elements), a certain function and a set of NFPs. All this information is modeled using the OWL2 format and stored in a triple store, which is a special type of database that is optimized for storing semantic data in triple format (subject-predicate-object). On execution, the workflow manager platform executes the defined process model with the related workflow engine. This workflow engine is also responsible for data management, i.e. to convey the needed data in the process, and handle exceptions. The logic of this engine is realized in a basic way (gets executed as defined), which could be extended with special routing or assistance functionalities in the future. For the execution of one task, related OWL2 instances from the BPD (modeled also with an ontology representation of BPMN 2.0 [13]) are provided for the Service Delivery Platform (SDP), which is responsible to allocate and call the right service. Several steps are needed, to find the proper web service which includes web service composition, when no proper atomic service may be found. A ranking score gets calculated during the steps, which is used to find the best service for invocation. The subsequent sections describe all steps in detail that are necessary to execute one complete task within a user-defined workflow. The task therefore describes a certain user need (defined as a goal) within a workflow, while the available services are defined as the building blocks that need to be assembled in order to fulfill the user's goal.

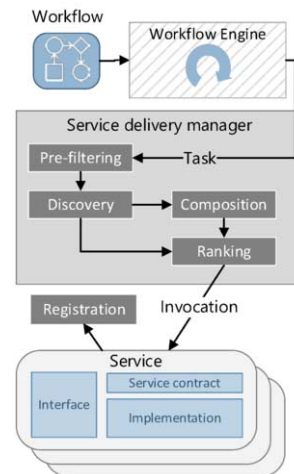


Figure 1 - Execution flow of an EHR workflow within the OntoHealth platform.

Step 0: Registration

Before pre-filtering, discovery, composition, ranking and invocation can take place, services need to be registered in the OntoHealth SDP. Registration is the process where the service descriptions are first generated by the service owner, according to the service contract, based on the semantic service model [14] and then stored in the triple store and managed by

² Tasks and services use the same domain ontology.

the SDP. These descriptions are given in JSON-LD³ format following the structure of the service contract. Each functionality of a service is defined by in-/outputs, function type and all corresponding NFPs. By means of SPARQL queries, the service descriptions can be added, updated or deleted from the registry. This step is not part of the actual workflow execution. In OntoHealth the definition of service information is task of the related developer, vendor or service provider. OntoHealth will provide a guideline on how to properly create the service description and register the service in the platform in order to make the service available for web service discovery.

Step 1: Pre-Filtering

The first step in the ranking procedure is the pre-filtering process. In this step, all the services that do not match the mandatory NFPs, which are retrieved from the task description, are dropped and not considered for further analysis. This helps to emphasize the discovery for services that require e.g. certain user permissions or any regional restriction that can be obtained from the requesting user. The resulting services are then used as an input for the following step.

It is defined in the semantic model if a NFP has the status of a mandatory property or will be considered later in the ranking step to calculate the ranking score.

Step 2: Discovery

The goal of the discovery step is to identify the services from all available services that fit the tasks' requirements defined by the user. The discovery of the services can be achieved based on functional classification and in- and output classification. Non-functional properties are considered in the ranking step (see Step 4).

Matching based on functional classification

After the pre-filtering is completed, the services have to be filtered according to the function they offer. Therefore the services are classified according to the following categories:

1. Full Match: Services in this group do match according to the function classification requested in the task and don't receive any penalty score during the ranking process.
2. Partial Match: Services in this group do not match the function classification requested in the task, but the function belongs to the same category (defined in the semantic model) as the one specified within the task. These services receive a penalty during ranking.
3. No Match: These services do not match according to the functional classification requested from the task. They are not considered during ranking.

The reason for defining those different categories is due to the further handling of each category in the further process.

Output Classification

In this part the services are compared with the related task and are classified based on their functional properties for the outputs. We distinguish between five different categories of matching as illustrated in figure 2:

1. Exact match: The output data elements of the task and the ones from the service match perfectly. There are no additional data elements delivered by the service.
2. Subsumption match: The service only delivers relevant output data elements, but not all of the functional requirements specified in the task are met, hence not all needed outputs are provided.

3. Plugin match: The service delivers all output data elements specified in the workflow task, but also delivers some additional ones.
4. Intersection match: The service only delivers some output data elements but not all of the ones specified in the task. Additionally, there are also additional outputs being delivered.
5. No match: The service delivers no relevant output data elements.

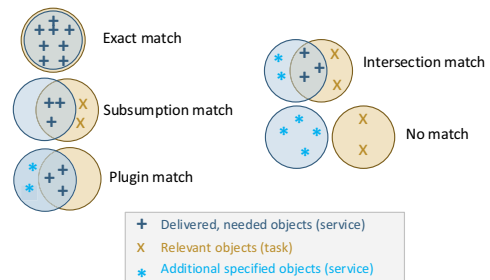


Figure 2 - Matching categories for output classification.

Input Classification

In this part, the services are classified based only on their functional properties based on matching the inputs of the task. We distinguish between two different categories of matching: match and no match:

1. Match: The services within this group match some or all input data elements specified within the task. This group is suitable for further service delivery consideration without additional processing.
2. No match: The needed inputs specified in the task are not fulfilled by this category. The services in this group require different data elements than the ones specified in the task and are considered for composition (see step 3).

All services, that are classified as subsumption, intersection or no match in at least one of the categories are not considered directly for the ranking step, but could be used in composition (see step 3). Related SPARQL-based queries consider dependencies of the functions, inputs, outputs and NFPs related to the semantic model. As all requirements are modeled using the developed WISE-DM ontology (see [14] for more details), inference could be used to get more precise results in pre-filtering and discovery.

In order to get more possibilities and freedom in requirements provision, a composition of services can be generated where outputs of service A can be used as the inputs for service B or both outputs can be combined to fulfill all the requirements specified within the task.

Step 3: Composition

After the matching from the discovery step has been completed, the resulting services are classified in the categories. While all services classified as full match for functions, exact- and plugin match for output classification and match for input classification do fulfill all functional properties without further effort, some additional processing is needed to probably make use of all other classifications. This is where composition comes in: all available services are arranged in groups where the outputs of service A plus all the already obtained outputs from other services fulfill the input of service B and so on.

³ <http://json-ld.org/>

The result of this service chain must then fulfill the functional properties of the related task. A composed service can then again be part of other composed services. The complete composition algorithm uses a graph-based structure as well as modified pathfinding algorithms. However, the detailed description would go beyond the scope of this paper.

A challenging part of composition is the NFP merging process, i.e. when combining the services to a group, also their NFPs have to be combined in a way that depends on the type of NFP. For example, for NFP “availability” the combined value is the lowest value of all involved services, while for NFP “price” the combined value is the sum of all involved services. The logic of the NFP considerations needs to be developed as a special ranking service that will be part of the next steps in the project. For the present conceptualization it is seen as a black-box or as a module with defined logic that follows a hardcoded pattern.

Step 4: Ranking

The ranking algorithm is responsible to find the best fitting service of all registered and composed services according to their NFPs. It takes as input the set of services (atomic or composed) and generates a list of ranked services ordered by their ranking score, which is defined as a unit-less number. During the ranking phase, all atomic or previously composed services get a ranking score based on how their NFPs fit compared to the ones specified in the task. The pseudocode of the algorithm can be found in code Table 1.

The actual score is not directly calculated from the NFP information (e.g. performance indicator), but is calculated via a distinct ranking service. This service is specified in the NFPs’ ranking instructions and is called during the ranking process. The arguments for the ranking service are all instances of an NFP from the task and the corresponding ones (same class) from the service. The idea of this ranking service is to define, what criteria makes a good or a bad score for a specified NFP, e.g. a ranking service for NFP costs determines, that a service with high costs should lead to a lower score. Each ranking service reveals a number as the percentage of suitability, which is then mapped to a score value for further calculations. The range of the score value is not predefined and can be adjusted.

During the workflow creation, the physician can specify an individual NFP weight depending on how important the NFP is considered on an individual base ($weight \geq 1$). An NFP adds its score multiplied by its weight to the score of the service. The function penalty score ($0 < penaltyScore \leq 1$) is integrated according to the functional classification (see step 2) to reduce the score in case of a partial match. The final score defines the suitability of the service. A higher score means, that more (or in the best case, all) NFPs are fulfilled by the service and thus increases the probability of getting selected for further processing.

Step 5: Invocation

The final phase in the service delivery process is the actual invocation of the service, which was previously discovered, ranked and selected according to the highest ranking score. Invocation enables the actual link between the semantic description of the service, according to the service model developed in earlier stages of the project and the actual service implementation. As we support various underlying implementations of services (e.g. REST, WSDL) the invocation would need to provide a generic interface for calling a service, but also be able to perform specific interaction calls, depending on the implementation of the service. Service usage including the service calls, related results and potential interaction will be

managed by the workflow manager. Those could be either used for further service calls, as a result presented to the user interface or any other process function defined within the workflow.

```

Input: Service S, Goal G
Output: Score of Service S

Ng = extracted NFPs of Goal G
Ns = extracted NFPs of Service S

score = 0;

for ns in Ns do {
    ng = getCorrespondingNfp(ns,Ng);
    tempScore = ng.weight * callRankingService(ns, ng);
}

score = tempScore * functionPenaltyScore;

return score;

```

Table 1 - Pseudocode of ranking algorithm. The “Goal” relates to the extracted workflow task.

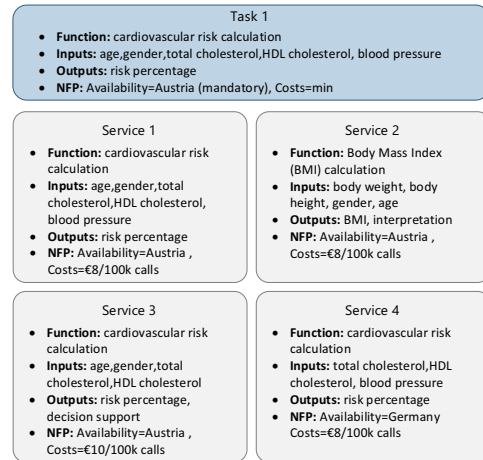


Figure 3 - Example Web Services. All information is semantically modeled and processed.

Service selection example

Let’s assume the physician wants to analyze the cardiovascular risk (*function*) considering the data “age”, “gender”, “total cholesterol (mg/Dl)”, “HDL cholesterol (mg/Dl)” and “systolic and diastolic blood pressure (mmHg)” (*inputs*) with a service highly available in Austria and with minimal costs (*NFPs*). The output is given through a percentage score value for “10-year risk of heart disease or stroke”. All this information is specified as a task and modeled semantically using our developed ontologies. The considered task and the available and registered services are depicted in figure 3. For demonstration purposes, the amount of NFPs has been limited to Availability and Costs. Although some of the services in this example share the same functionality, they differ in other relevant properties such as in-/outputs and NFPs.

Considering the risk calculation task and the available services in figure 3, Service 4 gets removed during pre-filtering, as the mandatory NFP is not supported. Service 2 does not fulfill the function, and Service 1 and 3 will be considered for the following ranking. Service 1 would relate to an exact match for output classification, Service 3 relates to a plugin match and both services are related to a match regarding the input classification. The ranking part would select Service 1 for final invocation, as the NFPs of this service lead to a better score

(considered NFP cost of Service 1 with €8/100k calls has a better score than Service 3 with €10/100k). No composition is needed for this example.

Implementation

The service delivery platform was developed as a Java application. Based on the semantic model from past project results, the RDF model was used with Apache JENA⁴ to model semantic dependencies during execution. The service delivery platform was designed according to the principle of a SOA. This means every step in the delivery process was realized by a single decoupled service. The services themselves can be seen as black-boxes. They were realized using the Jersey RESTful web services framework⁵ and communicate using JSON-LD. A manager web service was then introduced to facilitate the use of the platform by bundling service calls to the individual components to provide functionalities like, registering a new task or invoking a service based on a task.

Discussion

The developed Service Delivery Platform project is able to use proper web services according to the business process model of the user related workflow. The workflow engine processes the tasks and allocates proper web services based on input/output data elements, functions and non-functional properties. The discovery process got extended within this approach by using semantic technologies to find related suitable services (e.g. when a data element or function shares the same parent element, the parent or related siblings could be used as well) and composition, which uses semantic enriched service descriptions to compose several services in order to get the needed result. In contrast to approaches like FHIR, which is a service-oriented approach for accessing EHR data, the service delivery platform is responsible for the user-determined functionality in the sense of how accessible EHR data should be processed to extract, use and present the needed information related to the users' needs. The project SMART (Substitutable Medical Applications and Reusable Technologies) on FHIR [15] provides a platform for medical apps, that can be integrated into EHR systems based on individual user preferences. The main difference to our approach is, that SMART does not support the automatic discovery or composition of web services nor is it including non-functional properties for service discovery in order to reach the goals defined as a workflow. The service delivery platform as part of the ongoing project OntoHealth, will be refined based on future developments and will get its final evaluation at the end, when the complete software prototype will be tested and evaluated with real world scenarios and related end-users (physicians).

Conclusion

Providing means for users to define their information retrieval needs based on functional and non-functional properties at the point of care offers new ways of efficient EHR utilization. A workflow-based EHR-access is seen as a major advantage in efficient healthcare outcomes and even in user satisfaction. The next steps in the project will cover to enhance the usage of the semantic model for leveraging the service delivery process. This means, that related semantic reasons or related logic should provide better results on matching the requirements given from the workflow to the set of available services or

even use the information for more sophisticated and efficient composition.

Acknowledgements

This work has been supported by the Austrian Science Foundation (FWF), project number P 25895-B24.

References

- [1] Hsu W, Taira RK, El-Saden S, Kangaroo H, Bui A a T. Context-based electronic health record: Toward patient specific healthcare. *IEEE Trans Inf Technol Biomed* 2012;16:228–34.
- [2] Schweitzer M, Laserra N, Hoerbst A. Observing health professionals' workflow patterns for diabetes care - First steps towards an ontology for EHR services. *Stud Health Technol Inform* 2015;210:25–9.
- [3] Dolin R, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron P V, et al. HL7 clinical document architecture, release 2. *J Am Med Informatics Assoc* 2006;13:30–9.
- [4] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *Proc. 26th IEEE Int. Symp. Comput. Med. Syst., IEEE*; 2013, p. 326–31.
- [5] Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to Electronic Health Record Adoption: a Systematic Literature Review. *J Med Syst* 2016;40.
- [6] Unerl KM, Weinger MB, Johnson KB, Lorenzi NM. Describing and Modeling Workflow and Information Flow in Chronic Disease Care. *J Am Med Informatics Assoc* 2009;16:826–36.
- [7] Saleem JJ, Flanagan ME, Russ AL, McMullen CK, Elli L, Russell S a, et al. You and me and the computer makes three: variations in exam room use of the electronic health record. *J Am Med Inform Assoc* 2014;21:e147–51.
- [8] Raghupathi W, Kesh S. Interoperable Electronic Health Records Design: Towards a Service-Oriented Architecture. *E-Service J* 2007;5:39–57. doi:10.2979/ESJ.2007.5.3.39.
- [9] Gooch P, Roudsari A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Informatics Assoc* 2011;18:738–48.
- [10] Pavlovski CJ, Zou J. Non-Functional Requirements in Business Process Modeling. *Proc 5th Asia-Pacific Conf Concept Model* 2008;79:103–12.
- [11] Keller U, Lar R, Lausen H, Polleres A, Predoiu L, Toma I. Semantic Web Service Discovery. *Semant Web Serv* 2005:240–80.
- [12] Schweitzer M, Hörbst A. An Approach for the Support of Semantic Workflows in Electronic Health Records. "In Press 2017.
- [13] Rospocher M, Ghidini C, Serafini L. An ontology for the business process modelling notation. *Front Artif Intell Appl* 2014;267:133–46.
- [14] Laserra N, Schweitzer M, Gorfer T, Toma I, Hoerbst A. Building a semantic model to enhance the user's perceived functionality of the EHR. *Stud Health Technol Inform* 2016;228:137–41.
- [15] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc* 2016;1–10.

Address for correspondence

Marco Schweitzer
marco.schweitzer@umit.at
<https://ehealth.umit.at>

⁴ <https://jena.apache.org/>

⁵ <https://jersey.java.net/>

An OMOP CDM-Based Relational Database of Clinical Research Eligibility Criteria

Yuqi Si^a, Chunhua Weng^a

^aDepartment of Biomedical Informatics, Columbia University, New York, NY, USA

Abstract

Eligibility criteria are important for clinical research protocols or clinical practice guidelines for determining who qualify for studies and to whom clinical evidence is applicable, but the free-text format is not amenable for computational processing. In this paper, we described a practical method for transforming free-text clinical research eligibility criteria of Alzheimer's clinical trials into a structured relational database compliant with standards for medical terminologies and clinical data models. We utilized a hybrid natural language processing system and a concept normalization tool to extract medical terms in clinical research eligibility criteria and represent them using the OMOP Common Data Model (CDM) v5. We created a database schema design to store syntactic relations to facilitate efficient cohort queries. We further discussed the potential of applying this method to trials on other diseases and the promise of using it to accelerate clinical research with electronic health records.

Keywords:

Clinical Research Informatics; Relational Data Management; Electronic Health Record

Introduction

Randomized controlled trials are the gold standard for medical evidence generation. Eligibility criteria (EC) are the essential elements of clinical study protocols for specifying qualification of participants but often exist as free text, which are not amendable for computer processing. They are also found to have poor comprehensibility [1]. Given the wide adoption of electronic health records (EHRs), there is a great need for improving the interoperability of EC with EHRs to better integrate clinical research and patient care towards the development of a learning health system.

Multiple methods, such as ERGO [2], for structuring EC were developed before the widespread adoption of EHRs. As a result, such representations do not interoperate well with EHRs. Levi-Fix et al. developed EliXR-CDM [3] to structure criteria using the OMOP Common Data Model v4. This system was the first of its kind to transform free-text EC into a structured format using a standardized common data model. However, with a rule-based natural language processing system, it could not deal with the complex preprocessing and the scarcity of evaluation, which limited its generalization.

In this study, we extended this method and adopted the latest OMOP data standard, OMOP CDM version 5 [4], a model that is more comprehensive and better integrated than OMOP CDM version 4 for facilitating the interoperability among disparate observational databases. To the best of our knowledge, this study is amongst the first to build a relational database of clinical trial eligibility criteria using a widely adopted EHR data standard, OMOP CDM v5. Our method helps bridge the gap between clinical trials and EHRs by enabling fast and accurate

patient cohort searching for trial recruiters, protocol designers, and healthcare providers.

Method

Our method consists of the following steps: (1) criteria relational database design; (2) criteria parsing; (3) concept normalization using terminologies; (4) relation extraction; and (5) ETL (extract, transform and load) for criteria using the OMOP CDM v5. We used a hybrid machine learning-based natural language processing toolkit, CLAMP, for name entity recognition to extract medical terms in EC. We matched the extracted terms to the standardized concept identifiers in the OMOP CDM v5. Aside from the entity recognition, we also used the SVM classifier to obtain relations between entities and attributes. Finally, we built a relational database for fast querying via Django. We also provided a RESTful API for retrieving information.

Step 1: Database schema design

The EHR data standard of OMOP CDM v5 was described by the Observational Health Data Sciences and Informatics (OHDSI) community [5]. In this data model, medical terms were categorized into seven types including four entities (*Condition*, *Observation*, *Drug*, *Procedure*) and three attributes (*Qualifier*, *Measurements*, *Temporal_constraints*). Each attribute has a close relationship with a corresponding entity. For instance, a relation of *has_value* shows a quantitative measurement value of one entity. The four entities consist of medical terms with similar characteristics, while the three attributes differ from each other. Due to this, we decided to build an efficient schema in which the four entities could be stored into one table while the three attributes could be saved in three separate tables. The benefit of categorizing attributes in individual tables is to handle measurement and temporal constraints independently. This will prevent disarrangement with other terms in the criteria database as these two attributes are lab values or time phrases that need to be split in future work. We also used three types of relations to build the connections between entities and attributes. Given the fact that one entity has several attributes and one attribute corresponds to many entities, the relationship between entities and attributes were considered as many-to-many in the database. With this design, the relations could be saved in the database and the pattern of entities and attributes could be queried.

Step 2: Name entity recognition (NER)

To achieve precise name entity recognition, we implemented a comprehensive clinical natural language processing software, CLAMP [6], designed by Hua et al in 2015. We used annotated criteria corpus of 230 Alzheimer's disease clinical trial provided by previous lab members [7] to train the name entity recognition model. We implemented brown-clustering, n-gram, prefix-suffix, random-indexing, sentence-pattern, word-embedding, word-shape and word regular expression as name entity recognition features with a five-fold cross validation.

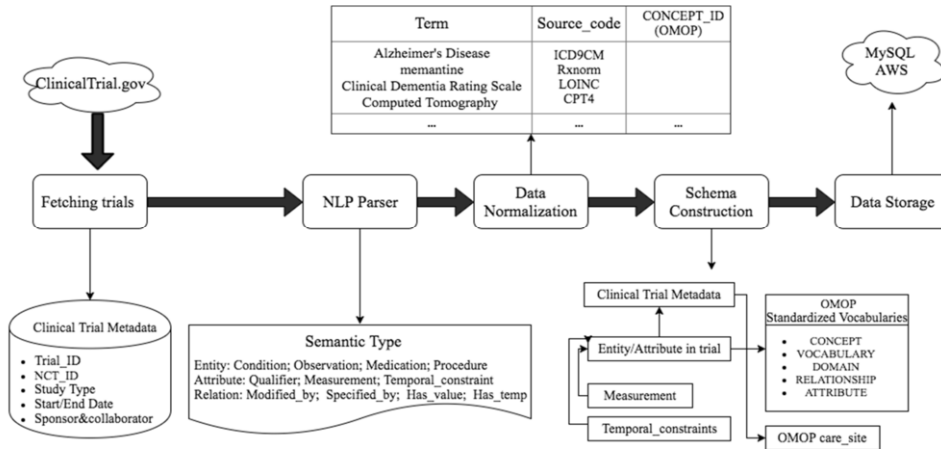


Figure 1– Workflow of Transformation and Reconstruction

Then we applied a NLP pipeline consisting of NegEx assertion, sentence detector, tokenizer, POS tagger, CRF-based NER and UMLS encoder. An example NER output is shown in Table 1.

Table 1– Structured Output of Entity and Attribute in EC

NCT00000171			
Exclusion Criteria:			
Sleep disturbance is acute (within the last 2 weeks).			
Condition	present	Sleep disturbance	T0
Qualifier	present	acute	T1
Temporal constraints	present	within the last 2 weeks	T2

Step 3: Concept normalization

Once we finished the name entity recognition, we mapped the extracted clinical terms into the concept standardization identifiers (CONCEPT_ID) using the open-source software, Usagi [8]. Each concept has a distinctive CONCEPT_ID, which is mapped to multiple CONCEPT_CODES across domains such as ICD9CM, SNOMED_CT, etc. With the matching, we were able to map the concepts in clinical research eligibility criteria into terminology standards. Usagi provided an algorithm to evaluate the effect of the matching by giving a score; a higher score represents better match, and a score of 1.00 is a 100% match. We manually reviewed 100 randomly chosen terms of each domain and analyzed a statistical performance of the matching score. After an assessment of the matching score, we set the matching threshold at 0.80.

Step 4: Relation extraction

We applied our previously developed open-source criteria parser [7] to extract relations between entities and attributes using the Support Vector Machine (SVM) classifier. The direction of each relation was defined from each entity to its corresponding attributes. This method used the basic function of LibSVM [9] with features including the class of head entity, the class of attribute, the shortest path between two terms in the dependency tree and whether or not the entity is the only one in its class in the corpus. The classifier inspected each entity-attribute pair and projected them into four classes: *no_relation*, *has_value*, *modified by*, *has_temp*. An example of relation extraction output is shown in Table 2. The relation between entities T4 (“liver or kidney disease”) and T3 (“clinically significant”) is “*modified by*”. The relation between entities T16 (“alcohol abuse and dependence”) and T15 (“current”) is “*has temporal relation*” or “*has-temp*” in short form.

Table 2– Structured Output of Relation in EC

NCT00007189		
Exclusion Criteria:		
Clinically significant liver or kidney disease.		
Current alcohol abuse or dependence.		
T4 liver or kidney disease	T3 Clinically significant	Modified by
T16 alcohol abuse or dependence	T15 Current	Has_temp

Step 5: Data storage

In the last step, we created an efficient schema using Django [10] and loaded all the extracted entities, attributes and relations into respective tables. The most economical method of storing relations is through many-to-many relationships. In addition, we used REST architecture [11] to build an API to provide a convenient interface for users to retrieve information.

Results

Database Infrastructure Description

The database is comprised of five major tables: (1) clinical trial metadata information (2) entity table (3) qualifier table (4) measurement table (5) temporal constraints table. The detailed schema and formulation of the database provided as an appendix is available at https://github.com/Yuqi92/DBMS_EC.

Descriptive Statistical Analysis

To understand how well the name entity recognition and relation extraction performs at each step, we designed an evaluation framework by using classical classification metrics: precision, recall and F-score, which are defined below: (TP: true positive; FP: false positive; FN: false negative; TN: true negative).

Table 3– Definition of TP, FP, FN, TN of a NER System

True positive	System extracts a concept that matches the label
False positive	System extracts a concept but there is no label or doesn't match the correct label

False negative	System doesn't extract a concept but there is a label
True negative	System doesn't extract a concept and there is no label

$$\text{precision} = \frac{TP}{TP + FP} \quad \text{recall} = \frac{TP}{TP + FN}$$

$$F1\text{score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

To ensure that we had an ample amount of training set for good performance on name entity recognition, we analyzed the performance of varying sizes of annotated files. Based on the learning curve shown in Figure 2, we confirmed that the training set of 230 annotated trials is sufficient to achieve good performance of name entity recognition.

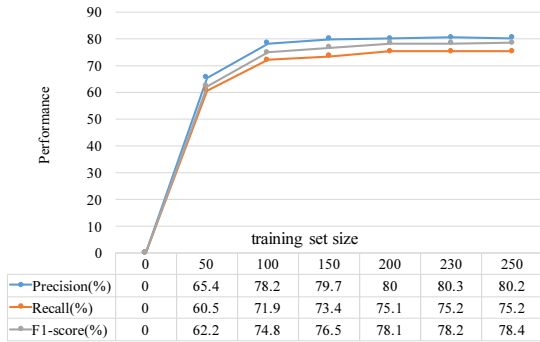


Figure 2— Learning curve for NER tasks

The performances of the different domains are variable due to the size of the training data. Of the annotated trials, the Condition domain is the largest (4136 terms) while the Procedure domain is the smallest (652 terms). According to the evaluation result of NER (Table 4), the Condition domain happened to achieve the best performance while the Procedure domain happened to have the poorest performance. Comparing these results, a larger amount correlates with a better performance and vice versa.

Table 4 – Evaluation of Name Entity Recognition

Domain	Precision	Recall	F1-score
Condition	0.835	0.836	0.831
Observation	0.748	0.745	0.793
Drug	0.852	0.790	0.820
Procedure	0.721	0.583	0.645
Qualifier	0.820	0.756	0.786
Measurement	0.820	0.770	0.794
Temporal constraints	0.826	0.788	0.807

We matched the four entities (*Condition*, *Observation*, *Drug*, *Procedure*) and the qualifier attribute to the CONCEPT_ID in OMOP CDM v5. We used the matching score to evaluate the mapping results for different domains on decreasing thresholds. Figure 3 is the descriptive statistical analysis curve of the matching score for the different domains. It's apparent and reasonable that when the matching score threshold was decreased from 0.9 to 0.7, the false positive rate (CONCEPT_ID incorrectly matched the term) decreased, while the false negative rate (CONCEPT_ID lost the term) increased. Therefore, we set the matching score threshold to 0.80 to trade off the balance between the error and the missing.

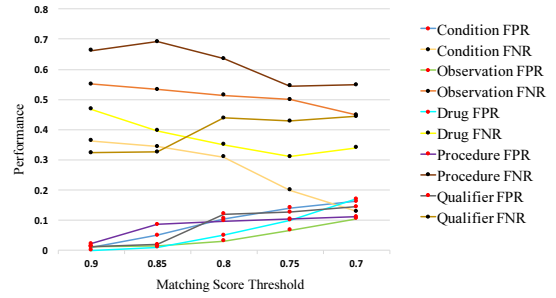


Figure 3— Mapping evaluation statistical analysis result

When the threshold reached 0.80, 68.80% of terms could be matched to CONCEPT_ID. Among the different domains, *Qualifier* terms reached the highest matching proportion (86.53%), *Condition* terms reached 74.88% and *Observation* terms reached only 44.41%. To evaluate the compression efficiency using CONCEPT_ID, we calculated the unique terms amount (in exacted terms), the compression ratio between number of unique terms and number of extracted terms, the unique CONCEPT_ID amount (in matched terms) and the compression ratio between number of unique CONCEPT_ID and number of matched terms. The ratio of the unique CONCEPT_ID is much lower than that of the unique terms in all domains. In total, the unique CONCEPT_ID ratio was 0.08, while the unique terms ratio was 0.19. Therefore, the representation ability of CONCEPT_ID was well proved. The details are shown in Table 5.

Table 5 – Statistical Matching Result of Extracted Terms

	Extra cted term	Num. of match	Perc. of match (%)	Unique term (compression ratio)	Unique CONCEPT_ID (compression ratio)
Condition	23336	17474	74.88	4453 (0.19)	1336 (0.08)
Observation	8824	3919	44.41	2360 (0.27)	391 (0.10)
Drug	6775	3694	54.52	1930 (0.28)	624 (0.17)
Procedure	3195	2136	66.85	626 (0.20)	193 (0.09)
Qualifier	9354	8094	86.53	449 (0.05)	188 (0.02)
Total	51484	35317	68.60	9819 (0.19)	2660 (0.08)

Relation extraction was evaluated separately by using the gold standard relations marked in annotated texts. The performance of SVM relation classifier is shown in Table 6. We counted the number of extracted relations and the number of attributes covered by the relations. Ideally, the attributes should not exist independently, and the cover percentage should be 100%. By dividing the number of corresponding attributes (*has_value* & *Measurement*; *has_temp* & *Temporal Constraints*; *modified by* & *Qualifier*), we calculated the percentage of extracted relations from the existing relations (Perc. of Extracting in Table 6). Our method extracted 54.81% of relations in general, 79.93% of relations between qualifier and entity, and 38.24% measurement.

Table 6 – Statistical Matching Result of Extracted Relations

	Number	Unique number	Attribute number	Perc. of Extracting (%)
Has_value	3005	2224	5816	38.24
Has_temp	4632	3051	4507	67.69

Modified by	10400	7477	9354	79.93
Total	18037	12752	19677	54.81

We also evaluated the relation extraction performance by manually reviewing 100 randomly-selected trials and counting statistical measurements including true positive, false positive and false negative. Then, we calculated the precision and recall of the three types of relation, “modified by”, “has temporal constraints” and “has value” as shown in Figure 4. The performance of relation “modified by” was the best among these three relations, while “has value” was the poorest. The performance evaluation result corresponds to the descriptive statistic matching result of extracted relations.

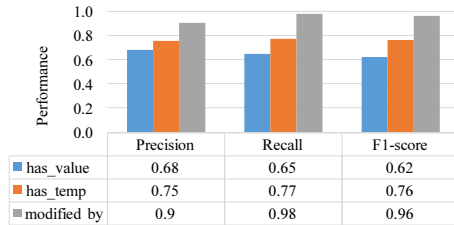


Figure 4— Evaluation of Relation Extraction

Sample Use Cases

This database of structured, standard-based eligibility criteria enables several use cases for integrating clinical research studies and electronic health records.

We have designed a RESTful API for users to input terms and search studies with certain criteria. Our http request accepted several parameters including entity key word (entity), criteria type (c-type), temporal constraint key word (t-constrain), qualifier (qua), etc. The sample request is in following format: GET {domain}/?entity=#&c_type=#&t_constrain=#&qua=#.

The response to this request is a list of clinical trials NCT identifiers that match the request.

Here we take several pair querying examples. For instance, if we are concerned about which trial has exclusion criteria involving participants with severe psychotic features within the previous three months, we are going to fetch all the available information for those specific parameters in the format {"entity": psychotic, "qualifier": severe, "criteria type": exclusion, "temporal constraints": three + months}. The complete URL for this request as it appears on the page: GET {domain}/?entity=psychotic&qualifier=severe&c_type=ex&t-constrain=three+month. The response to this request comes with two trials NCT identifiers: NCT01822951 and NCT00911807.

Another pairing example is to look for participants who have had stable AD therapy, which often occurs as essential inclusion criteria in Alzheimer’s disease clinical studies. In this case, the parameters should consist of entity (AD) and qualifier (stable). The following list of NCT identifiers is the response to this request:

NCT00495417, NCT02051608, NCT01122329, NCT02670083, NCT02386306, NCT01954550, NCT02423122, NCT00299988.

Therefore, by using the RESTful API, healthcare providers or clinical research investigators can request relevant study criteria information from our database.

Discussion

We fetched 1587 trials of Alzheimer’s disease as of September 2016 from ClinicalTrial.gov [12] and captured 4453, 2360, 1930, 626, 449 unique terms of *Condition*, *Observation*, *Drug*, *Procedure*, and *Qualifier* respectively. We also matched

extracted terms into CONCEPT_ID in OMOP CDM v5. The compression ratio of *Condition*, *Observation*, *Drug*, *Procedure*, and *Qualifier* were respectively 0.08, 0.10, 0.17, 0.09 and 0.02. Then we associated attributes with entities via relations including “has_value”, “has_temp” and “modified by” into a relational database. The relation “modified by” can be found and extracted from 79.99% of *Qualifier*. We justified the benefit of this method by descriptive statistical analysis and detailed user cases. We then further discussed the great potential and future application of this method in bridging the gap between EHR and EC.

Error Analysis

Errors of NER and relation extraction mainly resulted from wrong classified predictions. As for NER, the performance of the *Procedure* domain was poorer than that of other domains because the *Procedure* had the smallest number of instances in the training set. Since the output of NER is the input of the relation extraction, the errors in NER task will be multiplied in the relation extraction step.

Another cause of errors is the incomplete coverage of entities in the OMOP CDM v5. In other words, not all the terms existing in the criteria text have already been modeled in the OMOP CDM v5. Scarcity in the OMOP terminology dictionary is the reason why the matching score of some terms are lower than 0.50. Terms consisting of capital letters such as AChEI, NIA-AA criteria, MI are not identified correctly. An entire list of recommended terminology that could be added to the Concept table of OMOP CDM v5 will be provided.

Primary Contributions

This study has made four primary novel contributions.

First, we enabled semantic search of criteria by normalizing clinical terms using standard terminologies and by mapping them to CONCEPT_IDs in OMOP CDM. In this way, terms that share one meaning were regarded as the same. For instance, in the previous search methods, the term “AV block” was not returned by the query using the term “atrioventricular block”. In our database, these two terms are referred to one CONCEPT_ID, 316135. Users will no longer be inconvenienced by incomplete search results revolving around heterogeneous semantic representations for the same concept. Furthermore, each CONCEPT_ID has an associated clinical code such as ICD9CM, SNOMED_CT in the CONCEPT table of OMOP CDM. Users will be able to search for a specific disease by inputting its ICD9CM code.

The second primary contribution is the transformation of free-text criteria into a computable relational database compliant with an EHR common data model. The way relations were stored is a highlight of our work. For example, the many-to-many relationships in the database schema can retrieve relations between an entity and its respective attribute. Also, the clear definition and completeness of the attribute category will become a strong tool for handling pair querying, that the advanced search function provided by ClinicalTrial.gov could not achieve. For instance, if we input a combined search of several different domains such as “severe” + “Alzheimer’s Disease” + “for three years” + “inclusion”, then the search result will include all the trials with participants who have had severe Alzheimer’s disease for three years. Therefore, users can query and search the database for sophisticated logical queries, which can essentially improve the efficiency of clinical trial EC reuse.

Thirdly, the database of Alzheimer’s disease provides different audiences with an effective computer-based knowledge representation of EC. Study investigators could query in both the hospital data warehouse and the database of EC to target

eligible participants. Another use case for a trial designer is that the computable format of the criteria could help them define future study guidelines by comparing differences and commonalities of EC and study contents.

Finally, as an evidence-based clinical support method, the combination of searching the databases of EHR and EC allows healthcare providers to determine if a patient's treatment will benefit from a particular study or decide whether the patient is eligible for a study. Essentially, EHRs can be automatically matched to computably formatted clinical trial EC in our database. We could design a pipeline for patient screening with a combination of EHRs and the database of EC.

Limitations and Future Work

Based on the work we have done, researchers could build a database with more comprehensive information from clinical trial studies. Our future work will concentrate on two areas: performance and completeness. To improve the performance of our transformation pipeline for the free-text criteria, we will need to explore methods to extract complex expressions of *Temporal Constraints* [13] and *Measurement* [14]. We plan to extract specific numerical and temporal expressions from complex attributes. For example, temporal information such as “for three months” should be extracted and stored as “three” + “month” into different columns. Measurement information like “Hemoglobin $\geq 9.0\text{g/dL}$ ” should be extracted and stored by number and unit separately and the unit for the same test should be unified. Further collaborative research on natural language processing of free-text information is desired.

To improve the completeness, the outcomes and other sections of the trial will need to be transformed into structured output and stored into the database. We would also like to expand the method to cover the entire disease spectrum from ClinicalTrial.gov. More studies are warranted to test how this method would work for other eligibility features of other diseases. We may need to expand the database to better cover the eligibility features and elements. Additional tables may need to be added such as the *Anatomic Location* or *Genetic Name* when it comes to cancer. Furthermore, we will design a user-friendly interface to retrieve the necessary features from our database. The implementation of Django, a high-level web framework, also encourages rapid development and design that significantly reduces the workload of the back-end development. Therefore, we successfully transformed free-text EC into a computable, relational database following OMOP CDM v5. We hope that this computable format of EC can support the need of predictive analysis of targeted participants of clinical trials in the near future.

Conclusions

We contributed a practical method for transforming free-text eligibility criteria into a computable, relational database following OMOP Common Data Model (CDM) version 5. This method promises to be applicable to all disease trials in ClinicalTrial.gov and to accelerate EHR-based clinical research.

Acknowledgements

This research was supported by **R01 LM009886** from The National Library of Medicine (Bridging the semantic gap between research eligibility criteria and clinical data; PI:Weng).

References

- [1] Wu DT, Hanauer DA, Mei Q, Clark PM, An LC, Proulx J, Zeng QT, Vydiswaran VV, Collins-Thompson K, Zheng K. Assessing the readability of ClinicalTrials.gov. *Journal of the American Medical Informatics Association*. 2016 Mar 1; 23(2): 269-75.
- [2] Tu SW, Peleg M, Carini S, Bobak M, Ross J, Rubin D, Sim I. A practical method for transforming free-text eligibility criteria into computable criteria. *Journal of biomedical informatics*. 2011 Apr 30; 44(2): 239-50.
- [3] Levy-Fix G, Yaman A, Weng C. Structuring Clinical Trial Eligibility Criteria with Common Data Model. *Proceedings of 2015 AMIA Joint Summits for Translational Science*. San Francisco; 2015:194-198.
- [4] Observational Health Data Sciences and Informatics (OHDSI) OMOP Common Data Model V5.0. Available at <https://www.ohdsi.org>. Accessed Nov 2016.
- [5] Hripcsaka G, Dukeb JD, Shahe NH, Reichd CG, Huser V, Schuemief MJ, Suchardh MA, Parki RW, Wongf IC, Rijnbeekj PR. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Studies in health technology and informatics*. 216(2015): 574.
- [6] CLAMP: Clinical Language Annotation, Modeling and Processing Toolkit. Available at <http://clamp.uth.edu>. Accessed Nov 2016.
- [7] Kang T, Zhang S, Tang Y, Hruby GW, Rusanov A, Elhadad N, Weng C. EliE: An open-source information extraction system for clinical trial eligibility criteria. *Journal of the American Medical Informatics Association*. 2017 Apr 1. doi: 10.1093/jamia/ocx019.
- [8] Usagi, an application to help create mappings between coding systems and the Vocabulary standard concepts. Available at <https://github.com/OHDSI/usagi>. Accessed Nov 2016.
- [9] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2011 Apr 1;2 (3):27.
- [10] Django. Available at <https://www.djangoproject.com/>. Accessed Nov 2016.
- [11] RESTful Web services. Available at <http://www.restitutorial.com/>. Accessed Dec 2016.
- [12] ClinicalTrial.gov. <http://www.clinicaltrials.gov/>. Accessed Dec 2016.
- [13] Boland MR, Tu SW, Carini S, Sim I, Weng C. EliXR-TIME: a temporal knowledge representation for clinical research eligibility criteria. *AMIA Summits Transl Sci Proc*. 2012; 2012: 71-80.
- [14] Hao T, Liu H, Weng C. Valx: a system for extracting and structuring numeric lab test comparison statements from text. *Methods of information in medicine*. 2016; 55(3): 266-75.

Address for correspondence

Dr. Chunhua Weng, PhD, Department of Biomedical Informatics
622 W 168 Street, PH-20, New York, NY, 10032
Email: cw2384@cumc.columbia.edu

Characterizing Surgical Site Infection Signals in Clinical Notes

Steven J Skube^a, Zhen Hu^b, Elliot G Arsoniadis^{a,b}, Gyorgy J Simon^b,
Elizabeth C Wick^c, Clifford Y Ko^d, Genevieve B Melton^{a,b}

^a Department of Surgery, University of Minnesota, Minneapolis, MN, USA,

^b Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

^c Department of Surgery, University of California San Francisco, San Francisco, CA, USA

^d Department of Surgery, University of California Los Angeles, Los Angeles, CA, USA

Abstract

Surgical site infections (SSIs) are the most common and costly of hospital acquired infections. An important step in reducing SSIs is accurate SSI detection, which enables measurement and quality improvement, but currently remains expensive through manual chart review. Building off of previous work for automated and semi-automated SSI detection using expert-derived “strong features” from clinical notes, we hypothesized that additional SSI phrases may be contained in clinical notes. We systematically characterized phrases and expressions associated with SSIs. While 83% of expert-derived original terms overlapped with new terms and modifiers, an additional 362 modifiers associated with both positive and negative SSI signals were identified and 62 new base observations and actions were identified. Clinical note queries with the most common base terms revealed another 49 modifiers. Clinical notes contain a wide variety of expressions describing infections occurring among surgical specialties which may provide value in improving the performance of SSI detection algorithms.

Keywords:

Surgical Wound Infection; Quality and Safety; Text-mining

Introduction

Healthcare associated infections (HAIs) are a significant problem among hospitals worldwide. Surgical site infections (SSIs) are unfortunately the most common and costly of HAIs. SSIs increase post-operative morbidity and mortality. The overall prevalence of SSIs is estimated to be 2-5% for inpatient surgical cases in the United States [1-3]. Prevalence is significantly higher in certain specialties such as colon and rectal surgery (approximately 13-15%) [4].

An SSI can be classified into three categories (i.e., superficial, deep, and organ space) according to specific definitions which include time frame following surgery. Several classification systems designate an infection event as an SSI only if the occurrence is within 30 days of the index operation. A superficial SSI involves only the skin or subcutaneous tissue and requires documentation of one or more of the following: purulent drainage from the wound, aseptically obtained wound culture with isolated organisms, opening of the wound by a physician with clinical symptoms of infection, or diagnosis by a physician [5]. A deep SSI involves the muscle or fascia and requires: purulent drainage from the deep incision, opening of the wound, positive wound culture, or spontaneous dehiscing of the wound and clinical signs or symptoms of a wound infection, or an abscess or other evidence of infection diagnosed by pathology or imaging. An organ space SSI involves the space deep to the muscle or fascia, identification of involvement in a

specific organ system, and either purulent fluid from a organ space drain, an organism identified by culture, or an abscess or other evidence of infection diagnosed by pathology or imaging.

Another factor adding to the complexity of documenting SSIs is the question of a pre-existing infection. Careful documentation is required if there is an infection present at the time of surgery (PATOS). In the setting of PATOS, a post-operative infection would be excluded from being counted as an SSI if the site of infection is the same in both instances [5]. For example, if an abdominal wall abscess recurs after surgery performed for drainage of an abdominal wall abscess, this infection is not considered an SSI. However, if a patient undergoes surgery for intraabdominal abscesses from diverticulitis then develops a superficial infection of the wound, this is an SSI because the infection sites are different.

Many hospitals in the United States use the American College of Surgeons National Surgery Quality Improvement Program (NSQIP) as a quality improvement database to track post-operative complications such as SSIs. NSQIP is recognized as a national leader of post-operative complication measurement and quality improvement measures [6]. Currently, a surgical clinical reviewer who is usually a registered nurse employed by the hospital manually reviews the charts of post-operative patients and makes positive or negative SSI determinations by clinical judgement based on the SSI definition. This process results in high quality outcomes data for quality improvement and benchmarking efforts, but is prohibitively expensive for some centers [7]. Methods to automate or semi-automate SSI detection are of high interest since they may significantly reduce the burden of manual chart review and decrease the costs of quality improvement initiatives like NSQIP.

We previously developed supervised machine learning algorithms for SSI detection utilizing structured and unstructured clinical data [8]. SSI determination with our algorithms is based on a list of “strong features” identified for each type of SSI. The algorithm creates a score for each record correlating with the probability of an acquired SSI [9].

While the performance of our algorithms is good, improvements in the algorithm can be made particularly for intermediate scoring records and potentially in improving the feature set for unstructured text, which have to-date been based upon keywords and concepts derived from expert consensus (surgeons and hospital surgical clinical reviewers at our center). We hypothesized that there may be additional signals in the form of expressions directly describing or otherwise associated with SSIs in clinical notes. The study’s objective was therefore to characterize expressions associated with SSI determinations from clinical notes in a systematic manner.

Methods

Records from surgical patients included in the University of Minnesota Medical Center's NSQIP database were extracted from the University of Minnesota's clinical data repository. For our initial analysis, we included patients from 2014-2015 identified as having an SSI occurrence by the NSQIP surgical clinical reviewer along with patients having a high probability score (>40) for having an SSI from our SSI detection algorithm which used the following set of "strong" text features (Table 1). In all cases, the index operation was identified and all clinical notes within the 30 day time window after the operation were reviewed including all inpatient and ambulatory notes.

Table 1 – Original Surgical Site Infection Features

Surgical Site Infection Features			
abdominal abscess	empyema	antibiotics	phlegmon
abscess	erythema	joint abscess	presacral abscess
anastomotic dehiscence	evisceration	leak	purulent
cellulitis/cellulitic	extraluminal	interventional radiology	rectal stump blowout
cloudy	extravasation	malodorous	rim enhancing
dehiscence	fistula	murky	wet to dry
demarcated/demarcation	foul-smelling	open wound	wound dehiscence
drain care	Hartmann's blowout	packing/packing change	wound infection
drainage	induration/indurated	pelvic abscess	wound packing
drain placement	infected/infection	pelvic collection	vac dressing
dressing/dressing change	intra-abdominal abscess	pelvic sepsis	

In the review of each chart, one of two surgical residents (SS and EA) reviewed each post-operative note in chronological order from the index operation. All terms including those identified by experts previously within clinical notes contributory to an SSI determination were recorded along with information about misspellings, discrepancies and inaccuracies, and the associated note type containing the information. Documentation of repeated factors for an individual record was not recorded. All terms were carefully categorized into observations of the patient/patient data, actions performed by the clinical team, antibiotics, organisms, or clinical plans. These terms were also analyzed for modifiers and overall compared to the original set of terms.

Following categorization and initial analysis, we performed a validation of three "base terms" (i.e., "fluid collection", "drainage", and "infection") which were the most frequent and contained the greatest number of modifiers. From this analysis, we sought to validate the associated modifiers identified from our initial analysis. For this, we examined a separate cohort of patients with and without SSI (25 records each) from the year 2015. These were also patients within the the institution's NSQIP database, but had not been assessed for SSI-related phrases. We utilized the Natural Language Processing-Patient Information Extraction for Research (NLP-PIER) clinical research clinical note search engine for each base term [10]. Each of the encountered modifiers within the search engine were recorded and added to our representation model in our evaluation where applicable.

Institutional review board approval was obtained and informed consent waived for this minimal risk study. Interrater reliability was assessed on 160 (ten percent) of SSI phrases by both physician-raters (SS and EA) to assess agreement on the whether the phrase was associated with SSI and whether the term was positive or negative (i.e., not indicating an SSI). Agreement was 100% for association with SSI and 0.94 with a kappa of 0.82 for positive or negative designation agreement.

Results

A total of 54 positive SSI patient cases from the NSQIP database ($n=41$) or with a high probability SSI score by algorithm ($n=13$) were reviewed. After reaching 45 patients, saturation of our corpus was assessed by tracking new terms. After assessing 9 additional patients, only 8 new terms were identified and our cohort was completed with 54 total patients. Demographics and surgical specialty of these patient cases are displayed in Table 2. The total number of notes reviewed was 3,232. Multiple surgical services were represented.

Table 2 – Summary of 54 SSI patient cases

Demographics	
Median Age (Range)	55 (25-92)
Gender (%)	Male (48%)
Median Length of Stay in days (Range)	10 (1-43)
Median Number of Notes per chart (Range)	54(10-218)
Surgical Service	N (%)
General	11 (20%)
Colorectal	11 (20%)
Vascular	6 (11%)
Transplant	6 (11%)
Orthopedic	6 (11%)
Plastic	4 (7%)
Otolaryngology	4 (7%)
Neurologic	4 (7%)
Urology	2 (4%)

Overall, 1,536 distinct phrases were identified that were important for the designation of a SSI. There was a median of 25.5 (range 9-64) unique phrases identified in each chart. The majority of phrases 1,304 (85%) were identified from inpatient encounters. Outpatient encounters accounted for the remaining 15%, with 232 phrases. The majority of the SSI-related phrases were found in the progress notes of clinical teams following the patients in the hospital and is summarized in Table 3.

Table 3 – Location of SSI-Related Terms

Note Type	Terms	% of Total
History & Physical	78	5%
Operative Note	98	6%
Consultation	158	10%
Progress Note	869	57%
Primary Team	794	52%
Consult Team	75	5%
Discharge Summary	93	6%
Office Visit	119	8%
Telephone Encounter	45	3%
Emergency Visit	76	5%

While most of SSI-related phrases had a positive correlation with an SSI occurrence, 161 (10%) phrases offered evidence against a surgical site infection (e.g., "no obvious purulence", "improving of erythema", "wound c/d/i"). These phrases "protective" of an SSI generally occurred in the early post-

operative period, or late in the course of the infection, signifying potential recovery.

SSI-Related Base Term Classification

Observations were categorized by isolating the “base term” that was being observed. There were 63 unique base terms that were recorded from review of 1,536 SSI-related phrases (Table 4).

Table 4 - SSI Base Terms

Observation Base Terms		
incision	redness	fasciitis
debris	fluid collection	pain
skin	air	sepsis
succus	swelling	ecchymosis
drain	abscess	necrosis
induration	film	warmth
discharge	wound	tissue
output	purulence/pus	mesh
infection	seroma	fever
drainage	cellulitis	rigors
fluid	aspirate	edema
erythema	fistula	inflammation
dehiscence	thickening	blistering
leak	peritonitis	petechiae
hematoma	blood	purpura
odor	site	fluctuance
firmness	culture	separation
material	gas	osteomyelitis
dressings	stool	evisceration
tunneling	eschar	amputation
gangrene	exudate	colon

There were a few “actions” found to be pertinent to SSIs in clinical notes. Sixteen unique verbs were found relating to SSIs. Most verbs were found in multiple tenses. These verbs (Table 5) were documented when used to explain procedures relevant to surgical site infections.

Table 5 – SSI Related Actions

SSI Actions		
open	place	examine
incise	evacuate	close
remove	drain	culture
probe	washout	debride
aspirate	change	I&D
	irrigate	

Directives of the clinical plan also included some phrases related to SSIs. Each group of phrases had some variability, but there were six main themes: computed tomography (CT) requests, wound culture/gram stain orders, specific wound care plan, consulting infectious disease (ID) and interventional radiology (IR), tentative plans for operative intervention, and antibiotic changes.

Other SSI-Related Phrases

Antibiotics were common SSI-related phrases included in the clinical note. Antibiotics comprised 241 (16%) of the recorded SSI-related phrases. The use of antibiotics was not consistent in the treatment of SSIs. Antibiotics were included in the analysis if started empirically (concern for SSI but no definitive evidence) or if being used to treat an SSI. There was a wide range of antibiotics used for treating SSIs due to the multiple organ systems represented by each different surgical service. Documentation in the clinical notes included both the trade and generic names of the antibiotics (Table 6). General terms, such

as “antibiotics”, “IV antibiotics”, and “antibiosis” were also documented but are not included in Table 6.

Table 6 – SSI-Related Antibiotics

Generic Name (if used)	Trade Name (if used)	Common Abbreviations
amoxicillin-clavulanate	Augmentin	-
cefazolin	Ancef	-
piperacillin-tazobactam	Zosyn	pip-tazo
levofloxacin	Levaquin	-
metronidazole	Flagyl	-
ciprofloxacin	-	cipro
vancomycin	-	vanco, vanc
tigecycline	-	-
clindamycin	Cleocin	clinda
linezolid	Zyvox	-
mupirocin	Bactroban	-
ertapenem	Invanz	Erta
meropenem	Merrem	mero
cephalexin	Keflex	-
nafcillin	-	-
ampicillin-sulbactam	Unasyn	-
trimethoprim-sulfamethoxazole	Bactrim	TMP-SMX
ceftriaxone	Rocephin	CTX
micafungin	-	mica
fluconazole	Diflucan	-
minocycline	-	mino
doxycycline	-	doxy

Clinically significant organisms were documented in the clinical notes. These were recorded when associated with a wound or abscess culture (Table 7). Organisms comprised 111 (7%) of the total SSI-related terms. An organism isolated from a wound or abscess culture is sufficient on its own to diagnose an SSI [5]. Terms such as “gram positive cocci”, “lactose fermenting rods”, and “coagulase negative staphylococcus” were also recorded but not included in the table due to their generality in describing many infections.

Table 7 – SSI Related Organisms

Documented Organisms	
klebsiella	proteus
escherichia coli	staphylococcus
prevotella	pseudomonas
corynebacterium	pasteurella
streptococcus	anerococcus
achromobacter	veillonella
enterococcus	peptostreptococcus
bacteroides	clostridium

Wound care items were commonly encountered terms. Types of dressings including: gauze, Kerlix, NuGauze, Aquacel, wet to dry, and xeroform were frequently documented as SSI-related terms. The most commonly encountered wound care item was “wound vac”.

SSI-associated abbreviations and acronyms were relatively uncommon, but repeated frequently. The recorded abbreviations are included in Table 8. They are grouped according to type of abbreviation.

Table 8 – SSI Related Abbreviations

Category	Abbreviation
Anatomic Location	IT, LLQ, LUQ, RUQ, RLQ, EC, abd, RP
Microbiology	cx, abx, GM+, GPC, GNR, GPR, UTI, CoNS, ifx, MSSA, MRSA, VRE, VSE, GNB, SSI
Imaging	CT, A/P, US, IR
Exam	c/d/i, CDI, T _{max} , TTP
Laboratory	WBC
Frequency	BID, TID, QID
Miscellaneous	POD, op, IV, PO, s/s, I&D, ID, vac, JP

Misspellings in the clinical notes relating to surgical site infections were infrequent. There were only 12 instances of misspelling found in SSI related terms. The most commonly misspelled word was “dehiscence”. Inaccuracies of documentation associated with SSIs also appeared to be rare. Although it is difficult to assess inaccuracies by solely a retrospective search of the clinical notes, only one obvious inaccuracy was discovered. A pelvic abscess was incorrectly documented in a telephone note.

SSI Expressions in Notes Versus Original Expert Terms

As demonstrated in Table 1, there were 43 unique phrases determined by expert consensus that were included in the original set of “strong features”. These original phrases can be broken down into 22 base terms (observations and actions) and 24 modifiers.

Overall, extraction of SSI signals from clinical notes resulted in an overlap of 17 base observations and actions (77%) with the original set of expert phrases with 5 terms from the original list not found in our corpus. There were 24 modifiers identified from the original features, with an overlap of 21 modifiers (88%). Only 3 modifiers were not found in our corpus. Combined, there was 83% overlap when accounting for both base terms and modifiers. One term in the original set, “antibiotics”, was categorized in our other SSI-related features. In addition, we identified 62 new terms from the corpus: 47 new base observations and 15 new actions. Eleven of these terms were in the top 25% of frequency in the 54 cases reviewed. All of our antibiotics, organisms, and abbreviations/acronyms were new compared with the original set.

Modifiers of Base Terms & Validation of Modifiers

Modifiers of the main SSI-related base terms (observations and actions) were extracted in analysis of each base term. Modifiers were also classified as evidence for an SSI diagnosis (positive) and against an SSI diagnosis (negative). Overall, there were 383 modifiers among all of the base terms. Only unique modifiers were recorded for each base term. Repeated modifiers were recorded only if used for different base terms. There was a wide range of modifiers per base term, with a median of 2 (range 0-49).

Three terms: “fluid collection”, “drainage”, and “infection” were tested with the NLP-PIER search engine on a new set of 25 patients with an SSI and 25 patients without an SSI in 2015 to validate the utility of the previously identified modifiers and to determine if there were additional modifiers associated with these base terms. From chart review, “fluid collection”, “drainage”, and “infection” had the most modifiers, with 32, 49, and 41 respectively. Only 49 new modifiers were encountered in this clinical note query of 50 patients (13 for “fluid collection”, 18 for “drainage”, and 18 for “infection”). Figure 1 is an example of a base term with its modifiers.

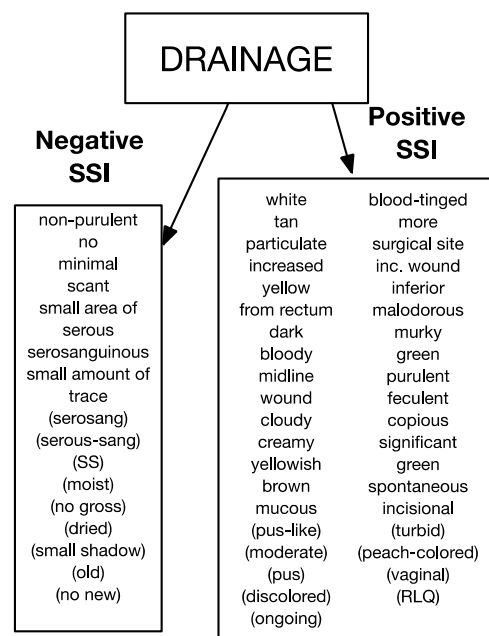


Figure 1 – Base Term Example and Modifiers. Additional modifiers identified by NLP-PIER are included in parentheses

Discussion

Automated or semi-automated SSI detection has the potential to decrease present day manual abstraction required in most cases. While quality improvement registries such as NSQIP demonstrate tangible benefits to patient outcomes in hospitals in the private and public sectors [7;11], increased automation around outcome extraction for post-operative complications like SSIs could reduce cost barriers creating wider adoption. Our study demonstrated a number of important types of signals in clinical texts which we did not recognize previously with the assistance of expert consensus. This speaks to the variability of language used in our documentation of patient care, and it is likely that these findings can be leveraged to improve the performance of these algorithms.

The method used in this study was a two-step approach by first empirically analyzing the content of clinical notes in positive cases of SSI and then performing a validation of the base terms and associated modifiers to ensure good coverage of the identified modifiers. A previous study reported defining SSI patterns using two conceptual groups of terms in text-mining: bacteriology and surgery [12]. In our study, we discovered additional groups of SSI-related terms. We opted to classify our concepts into observations, actions, and plans related to SSIs. Our SSI-related observations were used to describe exam, laboratory, and imaging findings. SSI-related actions were used for procedures and tasks performed by clinicians. The clinical plan includes future tasks to be performed and next steps in SSI management. Other information related to SSIs included antibiotics, organisms, and wound care. We also observed abbreviations for many of these concepts.

We observed that antibiotics and organisms comprised 23% of total terms. These terms appear to have a relatively high sensitivity in detecting SSIs since only certain organisms and antibiotics tend to be associated with post-operative infections. Unfortunately, these same antibiotics and organisms can often

be found in infections not relevant to SSIs, resulting in a low specificity for SSI detection.

The original set of “strong features” identified by expert opinion included a portion of terms that could be used for SSI identification. Compared to our original set, 62 new base terms and 362 new modifiers were identified in chart review of patients with an SSI or a high probability score on the SSI algorithm. In addition to the base terms, there were six clinical plan categories as well as antibiotics, organisms, and abbreviations/acronyms that could be leveraged for better SSI detection. While most modifiers were already discovered in the initial review of SSI cases, 49 additional new modifiers were identified through NLP-PIER searches which added to the robustness of the associated set of signals.

As expected, key data for the determination of SSIs is stored in the clinical notes. While structured data is useful for detecting SSIs, clinician judgement and physical examination remain key; significant, unique details about SSIs are only found within clinical notes. NLP and text-mining can be used to detect adverse events in clinical notes with better performance than manual review and other methods utilizing structured data for automated detection [13]. Recently, a text-mining approach using two categories was used to detect SSIs in a neurosurgery department [12]. Our findings demonstrate, however, that a constellation of terms is needed to determine the presence of an SSI. It is likely that improved discrimination for SSIs can be achieved by accounting for more complex phrases and/or base terms and their modifiers to capture more complex SSI semantics. Further organization including analysis of descriptor groupings and locations may be useful in classifying terms for the identification of SSIs.

This project did not assess pre-operative risk factors and predictors of SSIs in surgical patients. By studying post-operative outcomes and improving outcome abstraction in surgical patients, we are building our knowledge of these risk factors and predictors. These risk factors have even more complexity, ranging from patient physical characteristics, past medical and surgical history, additional laboratory data, to operation-specific signals. With more data analysis, perhaps a future project could assess pre-operative clinical notes to develop a separate algorithm to predict SSI risk prior to or immediately following an operation.

Our study has several limitations including its relatively small sample size and the use of data from a single institution. Future work to validate these findings on a separate dataset including any regional variability of SSI language and additional variability associated with specialties with lower rates of SSI (e.g. neurosurgery or otolaryngology) is needed.

Conclusions

The language behind SSIs is complex. There are many categories of terms that may contribute to an SSI determination, including observations, actions, clinical plans, antibiotics, and specific organisms. Empiric analysis of SSI cases was an effective method for uncovering the complexity of SSI-related expressions in clinical texts. These findings may provide value in improving the performance of SSI detection algorithms.

Acknowledgements

This research was supported by the University of Minnesota Academic Health Center Faculty Development Award (GS, GM), Agency for Healthcare Research and Quality (R01HS24532), National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program (UL1TR000114), Fairview Health Services, and University of Minnesota Physicians.

References

- [1] Anderson, D. J., Podgorny, K., Berrios-Torres, S. I., Bratzler, D. W., Dellinger, E. P., Greene, L., . . . Kaye, K. S. (2014). Strategies to prevent surgical site infections in acute care hospitals: 2014 update. *Infect Control Hosp Epidemiol*, 35 Suppl 2, S66-88.
- [2] Ban, K. A., Minei, J. P., Laronga, C., Harbrecht, B. G., Jensen, E. H., Fry, D. E., . . . Duane, T. M. (2016). American College of Surgeons and Surgical Infection Society: Surgical Site Infection Guidelines, 2016 Update. *J Am Coll Surg*. doi:10.1016/j.jamcollsurg.2016.10.029
- [3] Magill, S. S., Edwards, J. R., Bamberg, W., Beldavs, Z. G., Dumyati, G., Kainer, M. A., . . . Team, E. I. P. H.-A. I. a. A. U. P. S. (2014). Multistate point-prevalence survey of health care-associated infections. *N Engl J Med*, 370(13), 1198-1208. doi:10.1056/NEJMoa1306801
- [4] Wick, E. C., Vogel, J. D., Church, J. M., Remzi, F., & Fazio, V. W. (2009). Surgical site infections in a “high outlier” institution: are colorectal surgeons to blame? *Dis Colon Rectum*, 52(3), 374-379.
- [5] National NHS. Surgical Site Infection (SSI) Event. (2013) Atlanta, GA: Center for Disease Control and Prevention.
- [6] Ingraham, A. M., Richards, K. E., Hall, B. L., & Ko, C. Y. (2010). Quality improvement in surgery: the American College of Surgeons National Surgical Quality Improvement Program approach. *Adv Surg*, 44, 251-267.
- [7] Hollenbeak, C. S., Boltz, M. M., Wang, L., Schubart, J., Ortenzi, G., Zhu, J., & Dillon, P. W. (2011). Cost-effectiveness of the National Surgical Quality Improvement Program. *Ann Surg*, 254(4), 619-624.
- [8] Hu Z, Melton GB, Moeller ND, Arsoniadis EG, Wang Y, Kwaan MR, Jensen EH, Simon GJ. Accelerating Chart Review Using Automated Methods on Electronic Health Record Data for Postoperative Complications. *Proceedings of the American Medical Informatics Association Symposium*. 2016: 1822-31.
- [9] Hu, Z., Simon, G. J., Arsoniadis, E. G., Wang, Y., Kwaan, M. R., & Melton, G. B. (2015). Automated Detection of Postoperative Surgical Site Infections Using Supervised Methods with Electronic Health Record Data. *Stud Health Technol Inform*, 216, 706-710.
- [10] McEwan, R., Melton, G. B., Knoll, B. C., Wang, Y., Hultman, G., Dale, J. L., . . . Pakhomov, S. V. (2016). NLP-PIER: A Scalable Natural Language Processing, Indexing, and Searching
- [11] Dimick, J. B., Chen, S. L., Taheri, P. A., Henderson, W. G., Khuri, S. F., & Campbell, D. A. (2004). Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement Program. *J Am Coll Surg*, 199(4), 531-537.
- [12] Campillo-Gimenez, B., Garcelon, N., Jarno, P., Chaplain, J. M., & Cugia, M. (2013). Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France. *Stud Health Technol Inform*, 192, 572-575. doi:10.1016/j.jamcollsurg.2004.05.276
- [13] Melton, G. B., & Hripsak, G. (2005). Automated detection of adverse events using natural language processing of discharge summaries. *J Am Med Inform Assoc*, 12(4), 448-457. doi:10.1197/jamia.M1794

Address for correspondence

Genevieve B. Melton, MD, PhD
Institute for Health Informatics and Department of Surgery
420 Delaware Street SE, Mayo Mail Code 450
Minneapolis, MN 55455
Tel: + 612-625-7992
E-mail: gmelton@umn.edu

Formalizing Evidence Type Definitions for Drug-Drug Interaction Studies to Improve Evidence Base Curation

Joseph Utecht^a, Mathias Brochhausen^a, John Judkins^a, Jodi Schneider^b, Richard D. Boyce^c

^a Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA,

^b School of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, USA,

^c Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, USA

Abstract

In this research we aim to demonstrate that an ontology-based system can categorize potential drug-drug interaction (PDDI) evidence items into complex types based on a small set of simple questions. Such a method could increase the transparency and reliability of PDDI evidence evaluation, while also reducing the variations in content and seriousness ratings present in PDDI knowledge bases. We extended the DIDEO ontology with 44 formal evidence type definitions. We then manually annotated the evidence types of 30 evidence items. We tested an RDF/OWL representation of answers to a small number of simple questions about each of these 30 evidence items and showed that automatic inference can determine the detailed evidence types based on this small number of simpler questions. These results show proof-of-concept for a decision support infrastructure that frees the evidence evaluator from mastering relatively complex written evidence type definitions.

Keywords:

Drug Interactions; Ontologies; Artificial Intelligence

Introduction

While medication therapies are generally beneficial to a patient's health, they can also result in harm. A recent review of epidemiologic studies published over a 14 year period found that 3.5% of hospital admissions are the result of an adverse drug reaction [1]. The United States Department of Health and Human Services recently stated that reducing the rate of adverse drug events should be a national priority [2]. Research on the safety of any given drug starts during its development and continues after marketing. During these phases, studies examine the potential for one drug to alter the pharmacokinetic properties (absorption, distribution, metabolism, or distribution) or clinical effect (pharmacodynamics) of another drug. The results of these studies can suggest PDDIs that might lead to preventable harm to patients without proper management [3].

The spectrum of study types used to research PDDIs is broad and complex. These include in vitro experiments, population pharmacokinetic analyses, randomized controlled clinical trials, and observational epidemiologic studies [4]. Data mining adverse event reports and case report evaluation can also be included because these are research activities that generate PDDI hypotheses [5]. Not only are there many different types of PDDI studies, there are also numerous considerations that can influence the validity of a given study [2,4]. The range and complexity of study designs and the resulting evidence types can make it very difficult to evaluate a body of evidence to determine if a PDDI exists. This difficulty might be an

important factor influencing the large differences in content found in PDDI knowledge bases that are designed to help clinicians make management decisions [6,7]. Indeed, a more systematic approach to evaluating evidence for the existence of PDDIs was one of the recommendations put forth by a recent expert consensus conference series [8].

We believe that an ontology-based system can categorize PDDI evidence items into complex types based on a small set of simple questions, thereby reducing the cognitive load experienced by evidence evaluators. Our approach is to have a computer program infer the specific evidence type of a PDDI study based on simple data provided by the evidence evaluator. Here, we report our progress creating an infrastructure for this kind of decision support. We demonstrate a proof-of-concept for using an off the shelf OWL reasoner and formal evidence type definitions so that an evidence base curator only has to identify the study type at a high level (e.g., clinical study) and provide some simple design features (e.g., 'randomization'). With this information, the OWL reasoner can efficiently and consistently infer the specific evidence type with which to tag the study (see Fig. 1).

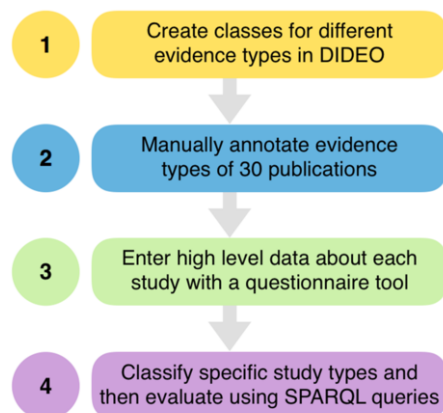


Figure 1 – The steps taken to build the proof-of-concept for categorizing PDDI evidence.

Methods

DIDEO (<https://github.com/DIDEO>) is an OWL ontology developed to facilitate managing information about PDDIs from multiple sources (clinical studies, case studies, in vitro experiments, etc.). It is freely accessible from <http://purl.obolibrary.org/obo/dideo.owl>. The fundamental ontological commitments of DIDEO have been described in detail [9]. One key contribution of DIDEO is to provide OWL classes for evidence types (see Fig. 2).

The process creating the textual definitions for each evidence type has been previously reported [10]. For Step 1 of the current project, we wrote necessary and sufficient conditions (i.e.,

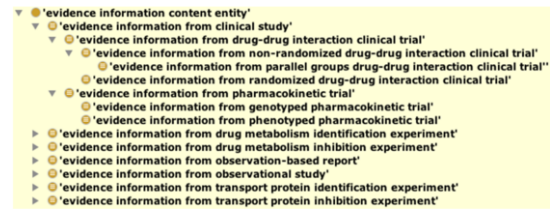


Figure 2 - The evidence information types from DIDEO. All subclasses of "evidence information from clinical study" currently existing in DIDEO are displayed.

Table 1 – (Step 1) Clinical trial evidence types represented in DIDEO, their textual definitions, and their necessary and sufficient conditions.

Evidence type (IRI and rdfs:label)	Definition (iao:definition)	Axiom
DIDEO_00000056 evidence information from clinical study	An evidence information content entity that is about a clinical drug trial.	'evidence information content entity' and ('is about' some (assay and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (('chemical entity' or 'drug product')))))
DIDEO_00000071 evidence information from drug-drug interaction clinical trial	An evidence information content entity that is about a clinical drug trial that has at least two drugs as its specified input.	'evidence information content entity' and ('is about' some ('scientific observation' and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'object drug role')))) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'precipitant drug role')))))
DIDEO_00000072 evidence information from non-randomized drug-drug interaction clinical trial	An evidence information content entity that is about a clinical drug trial that has at least two drugs as its specified input.	'evidence information content entity' and ('is about' some (assay and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (('chemical entity' or 'drug product') and ('bearer of' some 'object drug role')))) and (has_specified_input some (('chemical entity' or 'drug product') and ('bearer of' some 'precipitant drug role')))))
DIDEO_00000073 evidence information from parallel groups drug-drug interaction clinical trial'	An evidence information content entity that is about a clinical drug trial that has at least two drugs as its specified input, and that does not have group randomization as a part, and that realizes a clinical study design that has parallel group design as a part.	'evidence information content entity' and ('is about' some ('scientific observation' and (not (has_part value 'group randomization')) and (realizes some (concretizes some 'clinical study design' and (has_part some 'independent measure design')))) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'object drug role')))) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'precipitant drug role')))))
DIDEO_00000074 evidence information from randomized drug-drug interaction clinical trial	An evidence information content entity that is about a clinical drug trial that has at least two drugs as its specified input and does have group randomization as a part.	'evidence information content entity' and ('is about' some ('scientific observation' and (has_part some 'group randomization') and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'object drug role')))) and (has_specified_input some (('chemical entity' or 'drug product') and ('is bearer of' some 'precipitant drug role')))))
DIDEO_00000075 evidence information from pharmacokinetic trial	An evidence information content entity that is about a clinical drug trial that focusses on pharmacokinetics.	'evidence information content entity' and ('is about' some ('scientific observation' and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (organism and (participates_in some 'pharmacokinetic process')))) and (has_specified_input some ('chemical entity' or 'drug product')))))
DIDEO_00000076 evidence information from genotyped pharmacokinetic trial	An evidence information content entity that is about a clinical drug trial that focusses on pharmacokinetics and that has organisms as participants that participated in genotyping.	'evidence information content entity' and ('is about' some ('scientific observation' and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (organism and (participates_in some 'pharmacokinetic process') and (participates_in some genotyping))) and (has_specified_input some ('chemical entity' or 'drug product')))))
DIDEO_00000103 evidence information from phenotyped pharmacokinetic trial	An evidence information content entity that is about a clinical drug trial that focusses on pharmacokinetics and that has organisms as participants that participated in phenotyping.	'evidence information content entity' and ('is about' some ('scientific observation' and (realizes some (concretizes some 'clinical study design')) and (has_specified_input some (organism and (participates_in some 'pharmacokinetic process') and (participates_in some phenotype characterization))) and (has_specified_input some ('chemical entity' or 'drug product')))))

OWL equivalent class axioms) for each textual definition using entities defined or imported into DIDEO (see Table 1). We then ran an OWL reasoner (e.g. HermiT (<http://www.hermi-reasoner.com>)) to categorize the evidence types into a multi-level hierarchy.

Table 2. Five publications for each of the six clinical study evidence types defined in DIDEO

Evidence type by DIKB label	Publications (PubMed Identifier)
EV_PK_DDI_NR	10445377,10907965,15876900,11147928,8801057
EV_PK_DDI_Par_Grps	12911366,11910262,17571477,9855322,15518608
EV_PK_DDI_RCT	9542477,16778714,11563412,9757151,19242403
EV_CT_Pharmacokinetic	8911886,15834460,14747427,1487561,1438031
EV_CT_PK_Genotype	11452243,19142106,16765147,17429316,8689810
EV_CT_PK_Phenotype	8513845,7690693,2007317,1412613,8823236

Next for Step 2, we created a dataset that would enable automated decision support for people who need to evaluate PDDI evidence items. As Table 1 shows, evidence definitions for more specific kinds of evidence can become fairly complicated. Decision support should simplify the cognitive load of a person evaluating an evidence item.

We developed a proof-of-concept system that was used to test whether an OWL reasoner and formal evidence type definitions

can efficiently and consistently infer the specific evidence type from Table 1 based on basic evidence types (e.g., clinical study) and design features (e.g., ‘randomization’). Combining the evidence instance data with the OWL equivalent class axioms from DIDEO should enable an OWL reasoner to infer the specific evidence type of each instance. We tested this in Step 4 shown in Figure 1.

For Step 1 we used classes for the detailed evidence types shown in Table 1 in DIDEO (created in prior work) [11]. For Step 2, we drew on prior work [10] that had manually assigned a study type from Table 1: 30 publications were queried from the Drug Interaction Knowledge Base (<http://dikh.org/>), representing five publications for each of the six clinical study evidence types defined in DIDEO (see Table 2). Listing 1 shows the SPARQL query used to query a single evidence type from Table 2. The interested reader can modify the `dikhEvidence:Evidence_type` and PubMed identifier to retrieve

Listing 1. (Step 1) A query used to retrieve one of the 30 evidence types from the Drug Interaction Knowledge Base (<https://dbmi-icode-01.dbmi.pitt.edu/dikh/snorql>)

```
SELECT DISTINCT ?ev ?quote
WHERE {
  ?ev dikhEvidence:Evidence_type
      dikhEvidence:EV_PK_DDI_NR.
  ?ev rdfs:seeAlso
  <http://www.ncbi.nlm.nih.gov/pubmed/10445377>.
  ?ev siocns:content ?quote.
}
```

any of the other items listed.

As part of Step 3 a questionnaire tool, previously developed for the CAFE project (<https://cafe-trauma.com>), was modified and used to manually enter high level type and simple design features for each study in Table 2 (see Fig. 3). We then exported RDF instances created by the tool for each evidence item. We put these into a single file for import into an RDF store. Our RDF store was the community version of the Stardog (<http://stardog.com/>) RDF store, which comes with a built-in

DIDEO About Evidence List

Evidence Item Name: 15518608 [Next...] [RDF]

Trial

Method: Clinical Trial

Is there group randomization? ☐ Yes ☒ No

Is there parallel group design? ☒ Yes ☐ No

Did the study focus on pharmacokinetic processes? ☐ Yes ☒ No

Was phenotyping done? ☐ Yes ☒ No

Was genotyping done? ☐ Yes ☒ No

Drugs

pravastatin ☐ Object ☒ Precipitant

simvastatin ☐ Object ☒ Precipitant

atorvastatin ☐ Object ☒ Precipitant

verapamil ☒ Object ☐ Precipitant

Mibefradil ☒ Object ☐ Precipitant

Figure 3 – (Step 3) The questionnaire tool used to enter high level type and simple design features for each study in Table 2.

Table 3 – (Step 4) A set of competency questions and associated SPARQL queries that could be used by an application similar to that shown in Figure 3 to implement automated support to evidence evaluators for determining an evidence item's type.

Competency Question	SPARQL Query
Which evidence items stem from non-randomized clinical studies on drug-drug interactions without parallel group design?	<pre>select * { ?evidence a obo:DIDEO_00000072 . filter not exists { ?evidence a obo:DIDEO_00000073 . } }</pre>
Which evidence items stem from non-randomized clinical studies on drug-drug interactions with a parallel group design?	<pre>select * { ?evidence a obo:DIDEO_00000073 . }</pre>
Which evidence items stem from randomized clinical studies on drug-drug interactions?	<pre>select * { ?evidence a obo:DIDEO_00000074 . }</pre>
Which evidence items stem from clinical studies targeting pharmacokinetics without genotyping or phenotyping?	<pre>select * { ?evidence a obo:DIDEO_00000075 . filter not exists { ?evidence a obo:DIDEO_00000076 . } . filter not exists { ?evidence a obo:DIDEO_00000103 . } . }</pre>
Which evidence items stem from clinical studies targeting pharmacokinetics and using genotyping?	<pre>select * { ?evidence a obo:DIDEO_00000076 . }</pre>
Which evidence items stem from clinical studies targeting pharmacokinetics and using phenotyping?	<pre>select * { ?evidence a obo:DIDEO_00000103 . }</pre>

OWL description logic reasoner. We loaded RDF file of evidence instance data along with following OWL files:

- DIDEO: <http://purl.obolibrary.org/obo/dideo/release/2016-07-29/dideo.owl>
- RO core: <http://purl.obolibrary.org/obo/ro/releases/2016-09-02/core.owl>
- RO BFO classes minimal: <http://purl.obolibrary.org/obo/ro/releases/2016-09-02/bfo-classes-minimal.owl>
- IAO: <http://purl.obolibrary.org/obo/iao/2015-02-23/iao.owl>
- BFO (classes only) <http://purl.obolibrary.org/obo/bfo/2014-05-03/classes-only.owl>
- RO annotations: <http://purl.obolibrary.org/obo/ro/releases/2016-09-02/annotations.owl>

Finally for Step 4, we developed a set of competency questions and associated SPARQL queries (see Table 3) that could be implemented in an application similar to that shown in Figure 3 in order to provide automated support to evidence evaluators for determining an evidence item's type. We tested that each SPARQL query ran successfully over the inferred RDF store.

Results

Table 4 shows the results of running the competency question SPARQL queries (see Table 3) over the inferred RDF store. Each evidence type classification corresponds exactly to the manual classification from Table 2, thereby validating the data model and OWL ontology while showing proof of concept that complex evidence type classification can be obtained from the answers to simple questions.

Table 4 – (Step 4) Retrieval results from running the SPARQL queries (Table 3, Column B) on the Stardog triplestore.

DIDEO annotation	PMID of Publications
Evidence information from non-randomized DDI clinical trial (DIDEO_00000072)	10445377, 10907965, 15876900, 11147928, 8801057
Evidence information from parallel groups DDI (DIDEO_00000073)	12911366, 11910262, 17571477, 9855322, 15518608
Randomized drug-drug interaction clinical trial (DIDEO_00000074)	9542477, 16778714, 11563412, 9757151, 19242403
Pharmacokinetic trial (DIDEO_00000075)	8911886, 15834460, 14747427, 1487561, 1438031
Genotyped pharmacokinetic trial (DIDEO_00000076)	11452243, 19142106, 16765147, 17429316, 8689810
Phenotyped pharmacokinetic trial (DIDEO_00000103)	8513845, 7690693, 2007317, 1412613, 8823236

Discussion

DIDEO now provides OWL classes for many of the evidence types within the PDDI domain of scientific discourse. The current work shows the feasibility of using these formal definitions to build decision support that helps evidence

evaluators to determine the specific types of studies or experiments they review. Instead of having to master complex written evidence type definitions, such as those shown in Table 1, evaluators will only need to answer a small number of simple questions about the study, such as those shown Figure 3. With that information, simple SPARQL queries can be run over an inferred RDF dataset similar to the one used in this study to return the specific evidence type as a URI from DIDEO. By simplifying the task of evidence type assignment, a team of curators should be able to produce more correct and consistent work over the many hundreds of evidence items they need to manage.

Conclusions

In this work we showed proof-of-concept for the technical infrastructure, showing that PDDI evidence item data we added into a triple store together with DIDEO evidence types enable an OWL reasoner to infer an evidence item's specific type. In future work we plan to conduct a user study to test the hypothesis that this approach improves the correctness and consistency of evidence type assignment by evidence base curators. We also plan to extend the decision support to include questions that can be used to infer the quality of a study. We then plan to conduct a user study that compares interrater agreement between two groups of PDDI experts – one provided with evidence evaluation decision support, the other using their usual procedures. Based on our hypothesis, the group of participants given decision support should show much greater interrater agreement than the comparison group. We will also compare speed, ease, and user preference. If our hypothesis is shown to be true, it would be a promising step forward toward reducing the large differences in content found in PDDI knowledge bases that are designed to help clinicians make management decisions [6,7].

Evidence annotation is already a significant part of the biomedical enterprise, with computer-supported manual annotation used by an entire community and additional efforts to synthesize clinical research (e.g. Cochrane Reviews). We believe that approaches to simplify annotation processes, using incremental formalization and granular information, will be essential for increasing the availability of searchable (and in some cases algorithmically-synthesized) information. Generalizations of the work presented in this paper have the potential to greatly increase the impact of the curation enterprise.

Acknowledgements

This project was supported by a grant from the National Library of Medicine: “Addressing gaps in clinically useful evidence on drug-drug interactions” (R01LM011838). JS was partially supported by training grant T15LM007059 from the National Library of Medicine/National Institute of Dental and Craniofacial Research.

RDF descriptions for the 30 clinical studies generated in Step 3 are freely available from <https://goo.gl/Wk84gq>.

References

- [1] J.C. Bouvy, M.L.D. Bruin, M.A. Koopmanschap, Epidemiology of adverse drug reactions in Europe: A review of recent observational studies. *Drug Saf* **38** (2015), 437–453. US Department of Health and Human Services, *National Action Plan for Adverse Drug Event Prevention* [Internet]. 2014. Available from: <https://health.gov/hcq/pdfs/ade-action-plan-508c.pdf>

- [2] L. Magro, U. Moretti, R. Leone, Epidemiology and characteristics of adverse drug reactions caused by drug-drug interactions, *Expert Opin Drug Saf* **11** (2012), 83–94.
- [3] S. Hennessy, C.E. Leonard, J.J. Gagne, J.H. Flory, X. Han, C.M. Brensinger, W.B. Bilker, Pharmacoepidemiologic methods for studying the health effects of drug-drug interactions, *Clin Pharmacol Ther* **99** (2016), 92–100.
- [4] B. Percha, R.B. Altman, Informatics confronts drug-drug interactions, *Trends Pharmacol Sci* **34** (2013), 178–184.
- [5] L.M. Wang, M. Wong, J.M. Lightwood, C.M. Cheng, Black box warning contraindicated comedications: concordance among three major drug interaction screening programs, *Ann Pharmacother* **44** (2010), 28–34.
- [6] K.R. Saverno, L.E. Hines, T.L. Warholak, A.J. Grizzle, L. Babits, C. Clark, A.M. Taylor, D.C. Malone, Ability of pharmacy clinical decision-support software to alert users about clinically important drug–drug interactions, *J Am Med Inform Assoc* **18** (2011), 32–37.
- [7] R.T. Scheife, L.E. Hines, R.D. Boyce, S.P. Chung, J.D. Momper, C.D. Sommer, D.R. Abernethy, J.R. Horn, S.J. Sklar, S.K. Wong, G. Jones, M.L. Brown, A.J. Grizzle, S. Comes, T.L. Wilkins, C. Borst, M.A. Witte, D.C. Malone. Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support, *Drug Saf* **38** (2015), 197–206.
- [8] M. Brochhausen, J. Schneider, D. Malone, P. Empey, W.R. Hogan, R.D. Boyce, Towards a foundational representation of potential drug-drug interaction knowledge, *Proc. of DIKR 2014/IWOOD 2014/OBIB 2014 – Workshops of ICBO 2014*, Houston TX, USA, October 6–7, 2014, <http://ceur-ws.org/Vol-1309/paper2.pdf>.
- [9] R. Boyce, C. Collins, J. Horn, I. Kalet, Computing with evidence: Part I: A drug-mechanism evidence taxonomy oriented toward confidence assignment, *J Biomed Inform* **42**, 979–989.
- [10] M. Brochhausen, P.E. Empey, J. Schneider, W.R. Hogan, R.D. Boyce, Adding Evidence Type Representation to DIDEO, *Proc. Of the Joint International Conference on Biological Ontology and BioCreative* (2016), Corvallis OR, USA, August 1–4, 2016, http://ceur-ws.org/Vol-1747/IP02_ICBO2016.pdf.

Address for correspondence

Mathias Brochhausen: MBrochhausen@uams.edu
 Department of Biomedical Informatics
 University of Arkansas for Medical Sciences
 4301 W. Markham St., #782
 Little Rock, AR 72205-7199
 USA

A Method for Constructing a New Extensible Nomenclature for Clinical Coding Practices in Sub-Saharan Africa

Sven Van Laere, Marc Nyssen, Frank Verbeke

Department of Public Health (GEWE), Research Group of Biostatistics and Medical Informatics (BSI),
Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, 1090 Brussels, Belgium

Abstract

Clinical coding is a requirement to provide valuable data for billing, epidemiology and health care resource allocation. In sub-Saharan Africa, we observe a growing awareness of the need for coding of clinical data, not only in health insurances, but also in governments and the hospitals. Presently, coding systems in sub-Saharan Africa are often used for billing purposes. In this paper we consider the use of a nomenclature to also have a clinical impact. Often coding systems are assumed to be complex and too extensive to be used in daily practice. Here, we present a method for constructing a new nomenclature based on existing coding systems by considering a minimal subset in the sub-Saharan region. Evaluation of completeness will be done nationally using the requirements of national registries. The nomenclature requires an extension character for dealing with codes that have to be used for multiple registries. Hospitals will benefit most by using this extension character.

Keywords:

Clinical Coding; Names; Africa South of the Sahara

Introduction

Medical coding systems have shown their relevance in health information management (HIM) systems. Clinical coding is an integral part of HIM practice which provides valuable data for health care quality evaluation, health care resource allocation, health services research, medical billing, public health programming and Case-Mix/DRG funding.[1]

In sub-Saharan Africa, several countries use their own coding system by (1) setting up their own code system, (2) using an in-house designed code system based on another reference terminology, or (3) by making use of an existing code system. However, to the best of our knowledge, there is little literature about clinical coding in sub-Saharan Africa. The may be due to limited available capacity and resources to carry out comprehensive and significantly sound community assessments.[2]

Moreover, sub-Saharan countries often lack the structure and technologies to provide data. For this purpose, open-source hospital information systems such as OpenClinic GA [3] may help. A study in 19 African health facilities [4] demonstrated that by applying this software, structured reporting improved dramatically.

If we look towards interoperability between countries, still a lot can be done, since every health facility reports information in a different way. Moreover, the complex structures of existing coding systems limit the use of them. The idea of a new nomenclature for the sub-Saharan region that is extensible for a country or as more specialized hospital rise.

The word nomenclature has to be used with care in the context of sub-Saharan Africa. By the word “nomenclature”, usually they understand: “a coding system to provide codes that are used in invoices, to specify the health services that were provided for particular patients”. We define nomenclature in a broader way, also taking into account the clinical definition of health services.

In this paper we will present a way to build this new nomenclature for sub-Saharan Africa, based on existing nomenclatures.

Methods

In nomenclature construction, it is key to have a clear goal. For example, the existing Anatomical Therapeutic Chemical (ATC) classification is used as a tool to register drug utilization in order to improve the quality of drug use [5]. Our purpose is to see what is necessary per sub-Saharan country and to find a common subset between these countries (see Fig. 1).

We will start from the lists containing all health services that are necessary for billing purposes. This means we will at least include codes of existing procedure classifications. The Current Procedural Terminology (CPT), International Classification of Health Interventions (ICHI) and NOMESCO Classification for Surgical Procedures (NCSP) can help in providing these specifications.

However, not only the billing aspect of procedures are of interest in an electronic medical record. To provide more granularity compared to what is often used, we will also include codes for medication and lab results. Possible terminologies to include are respectively ATC and the Logical Observation Identifiers Names and Codes (LOINC). The idea is to design a new nomenclature of codes, at least including the

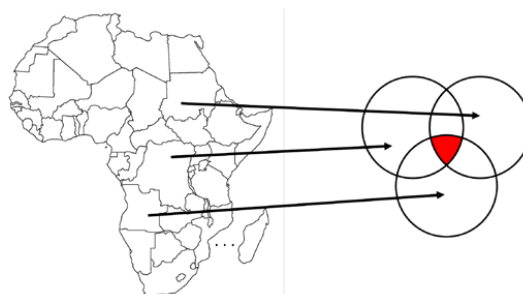


Figure 1: Idea of constructing a minimal subset of concepts (colored part) necessary to represent medical data

topics mentioned above, in which it is possible to refer to the original code used. To enable traceability, we will use following structure:

PREFIX.CODE[.XXX],

where *PREFIX* indicates a reference to the original terminology, *CODE* refers to the internal code of the code system that represents a clinical idea and an optional *XXX* part will refer to an index, necessary for capturing more granularity.

This last *XXX* part will enable the handling of codes that are not provided by the minimal subset, but that are necessary on a national or even more local (hospital) basis. This extensibility is one of the key aspects of this sub-Saharan nomenclature.

How do we know when this extension feature will be needed? For this evaluation on a national basis, we will make use of the information that is required to be provided by the hospitals for reporting purposes to the national registers. The data elements in the reporting requirements will be used as an evaluation source.

Three scenarios can occur in this evaluation part (see Fig. 2):

- **Scenario A:** If there is a 1-1 mapping from the newly constructed nomenclature to the national requirement, no problem will arise. Each new concept that is used, can be used to fill up the register regarding this concept.
- **Scenario B:** If there is a N-1 mapping from our new nomenclature to the national registry, it means two (or more) concepts will be used to fill up the national register data.
- **Scenario C:** If there is a 1-N mapping from the new sub-Saharan nomenclature towards the requirements of a national register, it is necessary to provide an extension. This extension will then split up the newly generated code into multiple parts using the *XXX* part, until we obtain at least one concept per national register requirement.

First we will evaluate this regarding national requirements. This extension of course can also be used on a more local basis, for example in an African hospital specialized for a particular health care service.

Nomenclatures

In this part we will briefly list up the code systems that can be of interest for the creation of this nomenclature. We need concepts dealing with procedures in which CPT, ICHI and NCSP. We also include the ATC and LOINC classifications for registering respectively medication and laboratory information.

ATC

ATC stands for Anatomical Therapeutic Chemical and is a classification system used for medication.[5] It classifies substances based on the organ or system on which they operate and their therapeutic, pharmacological and chemical properties and is maintained by the World Health Organization (WHO). Moreover, the ATC system also includes “defined daily doses” (DDDs) for many drugs, indicating the usual dose used per day. The classification uses the representation of the concept identifier to classify medication at five different levels: (1) anatomic main group; (2) therapeutic subgroup; (3) pharmacological subgroup; (4) chemical subgroup and (5) chemical substance.

CPT

The Current Procedural Terminology (CPT) is a medical code set for uniform communication about services and procedures specifying which services health care providers have performed on the patient.[6] This terminology is maintained by

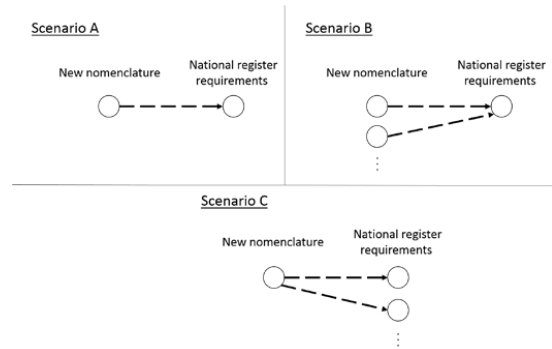


Figure 2: Possible mapping scenarios between newly composed nomenclature (left) and national register requirements (right);

Scenario A: mapping 1-1 (no problem)

Scenario B: mapping N-1 (no problem)

Scenario C: mapping 1-N (extension required)

the American Medical Association through the CPT Editorial Panel. Governments can use CPT for tracking the prevalence and value of procedures. Each CPT code is 5 characters long and may be numeric or alphanumeric, depending on which category the CPT code is classified. CPT codes are divided into 3 categories based on their usage.

ICHI Alpha 2015

The International Classification of Health Interventions (ICHI) is a system of classifying procedure codes, maintained by WHO.[7] ICHI was originally designed to replace the International Classification of Procedures in Medicine (ICPM). The classification is built around three axes: the *target* (the entity on which the Action is carried out), the performed *action* (a deed performed by an actor to a target) and the *means* (the processes and methods by which the ‘action’ is carried out). Extension codes are provided to allow users to describe more detail about the intervention in addition to the relevant ICHI code.

LOINC

The Logical Observation Identifiers Names and Codes (LOINC) provides universal identifiers for laboratory and other clinical observations.[8] This classification is maintained by the Regenstrief Institute. LOINC codes are composed by 5 or 6 parts that are separated by a colon: (1) component/analyte, (2) property observed, (3) timing of measurement, (4) type of sample, (5) scale of measurement, and if relevant (6) the method of the measurement.

NCSP

The NOMESCO Classification of Surgical Procedures (NCSP) is a classification for surgical procedures, maintained by Nordic Centre for Classifications in Health Care.[9] It is based on the traditions of the surgical profession in the Nordic countries. A NCSP code consists of three alphabetic characters (positions 1-3 of the code) and two numeric characters (positions 4-5 of the code). The alphanumeric part refers the classification (i.e. chapter) to which the code belongs and the numeric part refers to the specific procedure within the procedure group, including surgical technique and precise anatomical location.

Discussion

In sub-Saharan Africa many actors are willing to improve the health care system, especially the documentation of it. Not

only health insurance companies, but also ministries, public and private hospitals are ready to improve documentation. Nowadays most of the health care actions still are documented on paper patient charts or remain (partly) undocumented. When documentation is provided on these charts or in health information systems (HIS), registration is often incomplete, inaccurate and/or performed in an untimely way.[10]

The reason for this, is that documentation schemes are just too complex and too time-consuming (in the opinion of the management of a hospital and for the nursing staff). However, a mental change in sub-Saharan Africa is progressing: they tend to see benefits of documenting.

In HIS used in sub-Saharan countries, documentation is often used for billing purposes where each intervention that requires several sub-actions is summarized into one general concept. For example, the documentation of the concept "child-birth" does not only include the hospitalization, but also procedures and health services by various caregivers to help the patient in delivering a baby. However, in most cases hospitals will provide an invoice only based on giving child birth.

It is important to stress that the documentation of these various actions will lead to possible new insights regarding the functioning of the hospital and/or the country's health system. Moreover, by documenting every step of the care process, possible improvements with beneficiary outcome for the country and/or hospital are possible. National institutes will be able to provide epidemiologic summaries enabling them to improve the health status of their country. Hospitals will be able to calculate cost-benefit ratios based on what is documented and may adapt their price system on their findings.

The aim of this paper is introduce a method for creating a new extensible documentation system based on existing nomenclatures for the sub-Saharan part of Africa. The reason is an increasing awareness for the need of proper coding principles. Contrary to introducing a completely new coding system, we will use existing coding systems. The benefit of reusing existing terminologies is that this terminology will require less maintenance compared to a completely new terminology.

We will construct this nomenclature by calculating a common subset of concepts over the sub-Saharan countries. After this, an evaluation study is required to see if the new nomenclature covers every aspect necessary. The idea of extensibility is essential to solve the problems cause by the different national registry requirements in every country. To meet these requirements it is possibly necessary to extend the new nomenclature so that codes that map onto multiple register requirements can be split up, in order to be more fine-grained.

Not only nation-wide this extensibility will give improvements, but also hospitals can benefit from this characteristic of this new nomenclature. Sometimes procedures are only executed in particular hospitals. These procedure codes will then not be part of the general nomenclature, but will be put in a local extension.

Finally, we will enable interoperability between countries using this nomenclature. The power of this terminology lies in the origin of the common subset including only codes that are necessary.

Conclusion

Sub-Saharan Africa is willing to improve the health care systems by investing in good documentation practices. This increasing awareness grows both in nationwide institutions, as well as inside the hospitals. Current terminology systems appear to be complex and too extensive in use. We propose a method for constructing a new nomenclature based upon exist-

ing terminologies. To cope with national and local differences in sub-Saharan regions, this new nomenclature is made extensible. For evaluating this nomenclature we will use the registry information per country. When one code is necessary for providing the required information for multiple national registries, an extension will be the solution for providing the country with the correct data. This nomenclature has the benefit of being adaptable to both the needs of the hospitals and to the needs of the countries.

References

- [1] I. T. Adeleke, O. O. Ajayi, A. B. Jimoh, A. A. Adebisi, S. A. Omokanye, and M. K. Jegede, Current clinical coding practices and implementation of ICD-10 in Africa: A survey of Nigerian hospitals, *Am J Health Res* 3 (2015), 38-46.
- [2] D. T. Jamison, R.G. Feachem, M.W. Makgoba, E.R. Bos, F.K. Baingana, K.J. Hofman, and K.O. Rogo, *Disease and Mortality in Sub-Saharan Africa, 2nd edition*, The World Bank, Washington, NW, 2006.
- [3] F. Verbeke, OpenClinic GA [Internet]. Available from: <https://sourceforge.net/projects/open-clinic/> [cited 2016 Nov 29].
- [4] F. Verbeke, G. Karara, and M. Nyssen, Evaluating the impact of ICT-tools on health care delivery in sub-Saharan hospitals, *Stud Health Technol Inf* 192 (2013), 520-523.
- [5] WHO Collaborating Centre for Drug Statistics Methodology. Structure and Principles (of ATC) [Internet]. Available from: http://www.whocc.no/atc/structure_and_principles/ [cited 2016 Nov 29].
- [6] American Medical Association, Current Procedural Terminology (CPT®) Editorial Panel [Internet]. Available from: <https://www.ama-assn.org/about-us/current-procedural-terminology-cpt-editorial-panel> [cited 2016 Nov 29].
- [7] World Health Organization, International Classification of Health Interventions (ICHI) [Internet]. Available from: <http://www.who.int/classifications/ichi/en/> [cited 2016 Nov 29].
- [8] Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee, Logical Observation Identifiers Names and Codes (LOINC®) - Users' Guide [Internet]. Available from: <http://loinc.org/downloads/files/LOINCManual.pdf> [cited 2016 Nov 29].
- [9] Nordic Medico-Statistical Committee (NOMESCO), NOMESCO Classification of Surgical Procedures Version 1.16 [Internet]. Available from: http://www.nordclass.se/NCSP_1_16.pdf. [cited 2016 Nov 29].
- [10] W. Mutale, N. Chintu, C. Amoroso, K. Awoonor-Williams, J. Phillips, et. al., Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative, *BMC Health Serv Res* 13 (Suppl 2) (2013), S9.

Address for correspondence

Sven Van Laere,
Vrije Universiteit Brussel, Department of Public Health,
Research group of Medical Informatics and Biostatistics
Laarbeeklaan 103, B-1090 Brussels, Belgium
svvlaere@vub.ac.be

Content and Trends in Medical Informatics Publications over the Past Two Decades

Liqin Wang^{a,b}, Maxim Topaz^a, Joseph M. Plasek^c, Li Zhou^{a,b,d}

^a Division of General Medicine and Primary Care, Brigham and Women's Hospital, Boston, MA, USA,

^b Harvard Medical School, Boston, MA, USA

^c Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA

^d Clinical Informatics, Partners eCare, Partners HealthCare System, Boston, MA, USA

Abstract

This study aims to identify subject content and trends in the medical informatics literature in order to shed light on the past, current, and future directions of diverse education and research activities. A list of 36 core medical informatics journals was compiled through expert consensus. We retrieved 60,862 articles from the U.S. National Library of Medicine's MEDLINE database that were published by those journals from 1992 to 2015. A series of descriptive analyses were conducted to reveal the historical productivity of the journals, publications trends, and the subject content based on the Medical Subject Headings (MeSH) term frequencies and debut years. We found 73 core topics and 72 new topics of medical informatics within three relevant MeSH categories (informatics, techniques, and healthcare).

Keywords:

Medical Informatics; Publications; Medical Subject Headings

Introduction

Medical informatics (MI) is defined by the United States (U.S.) National Library of Medicine (NLM) as “the interdisciplinary study of the design, development, adoption, and application of IT-based innovations in healthcare services delivery, management, and planning” [1]. It emerged in the 1950s and has been an interdisciplinary domain for over half a century, involving applications of computer science, information science, engineering, social science, management science, among many others, in all fields of health and medicine [2]. Clinical informatics is now a medical subspecialty that allows physicians to practice clinical informatics in collaboration with other healthcare and information technology professionals to promote safe, efficient, and effective patient-centered care [3]. Today, over 80 U.S.-based academic institutions offer onsite or distance MI training programs [4].

Most published review articles in the MI domain have focused on specific subdomains or research topics, such as clinical decision support, natural language processing, and terminologies. Only a few prior studies have attempted to characterize the overall research activities in the MI literature [5]. To summarize the literature trends, these studies used Medical Subject Headings (MeSH), a comprehensive controlled vocabulary for indexing journal articles and books in the life sciences developed by NLM [6]. MeSH terms are assigned manually by professional indexers and the process can be assisted with automated approaches [7]. DeShazo et al. took the first step to use MeSH terms to analyze publication trends in the MI literature published between 1987 and 2006 [5]. Their study included all articles assigned any term in the

MeSH “Medical Informatics” hierarchy as a major topic across all types of journals indexed by MEDLINE/PubMed. To our best knowledge, no literature trend analysis across the overall MI field in the recent decade has been published.

The focus of this paper is to describe the evolution of the field of MI in the recent two decades by conducting a trend analysis on research topics published in MI journals using MeSH descriptor terms. The aim of this study is to update the MI community about the most recent research trends from published MI literature. The results of this study will provide important insights for future educational, research and development directions. For example, our findings may be used to aid in the curriculum development for the MI training programs, core content updates for the MI subspecialty exam [8], as well as suggesting potential research funding areas.

Methods

We utilized a multi-step approach to identify informatics publications and research topic trends: (1) identify core MI journals, (2) form a query to retrieve articles from the MEDLINE database, (3) extract MeSH terms from retrieved articles, (4) select MeSH term categories, and (5) perform descriptive analysis to identify trends (see Figure 1).



Figure 1—Methods Overview

Step 1: Identifying MI journals

In the MEDLINE database, there are approximately 26,000 journal records [9]. To identify MI publications, DeShazo searched MEDLINE for all documents assigned the term “Medical Informatics” in MeSH or any term in the “Medical Informatics” hierarchy as a Major Topic [5]. However, we found that this approach produces somewhat noisy results with a significant amount of non-relevant articles. For example, “databases, factual” is a descendant term in the MeSH Medical Informatics hierarchy. When one uses this term as a MeSH Major Topic to search MEDLINE, more than 29,000 articles are retrieved. However, many studies such as clinical trials, epidemiological studies, or economic analyses that used some sorts of databases may not necessarily be classified as “true” MI articles. Additionally, we also reviewed the 1200+ articles retrieved from the journal of “BMC Genomics” based on DeShazo’s query, the majority of the articles were not about MI studies per se, as this journal mainly focuses on genome-scale analysis, functional genomics, and proteomics. To account for this, we utilized a different approach by identifying MI “core” journals first. We

assume that the majority of MI topics are published in the “core” MI journals and topic trends in MI journals can serve as a proxy for the MI topic trends across journals in other fields.

In this study, we specifically include journals and articles related to clinical informatics, public health informatics, and nursing informatics that directly associate with patient care and population health. Although the broad biomedical informatics includes other important branches, such as bioinformatics and imaging informatics, journals focusing on these branches are not included in this analysis. For the purposes of this study, to define MI core journals, we examined whether the majority of the articles in a journal reflect the core content of clinical informatics as defined by Gardner *et al.* [8]. Some journals that cover those topics to a certain extent, but mainly fall into other domains, were excluded. These domains include: (1) biomedical engineering, physics, devices, images; (2) bioinformatics (except journal names containing ‘biomedical informatics’); and (3) general medical, nursing, and pharmacy journals.

We defined the inclusion and exclusion criteria and identified the core MI journals in two sub-steps: First, we started with a simple query to the MEDLINE/PubMed in November 2016 for retrieving all documents assigned any MeSH term in the following three MeSH hierarchies: “Medical Informatics”, “Public Health Informatics”, and “Nursing Informatics”. This query retrieved over 379,000 articles that were published in over 9,000 journals indexed by MEDLINE. The publication sources were sorted by the number of retrieved articles. Two MI researchers (LZ, MT) manually reviewed all journals with 50 or more relevant articles. For journals with less than 50 articles, or approximately 90% of publication sources, they used keyword search (e.g., informatics, information technology, and telemedicine) to identify MI journals. The two researchers separately reviewed the journal list and determined whether a source should be included or not. Disagreements were discussed and full consensus on MI journals was achieved. Second, we generated a list of MI journals by searching the NLM Catalog with the query “Medical Informatics [Broad Subject Term]”, which returned 91 journals indexed in MEDLINE. We then manually reviewed these 91 journals to identify MI journals. We found that many journals changed their names over the years. We merged different versions of journal names with the preferred name being the most recent. Finally, we compared and reconciled the journal lists generated by these two substeps.

Steps 2: MI Article Query

We formed a MEDLINE query based upon the identified core MI journals to retrieve all articles published in these journals between 1992 and 2015, as follows: *MI Journals AND "english" [language] AND 1992[PDAT]:2015[PDAT]*. In this query, the *MI_Journals* refer to the journals from the previous step, where both the current and previous journal names were used. This expanded our query format to include the variations: *("Stud Health Technol Inform" [Journal] OR "AMIA Annu Symp Proc"[Journal] OR "Methods Inf Med" [Journal]...)*. With this query, we were able to plot the number of MI articles over the past 24 years to see the increasing rate of publications in the MI field.

Step 3: MeSH Term and Article Information Extraction

All of the information about the retrieved articles was stored in MEDLINE’s Extensible Markup Language (XML) formatted files. We downloaded related files through NLM e-utilities and parsed those XML files using the Java DOM Parser to get each article’s identifying number (PMID), title,

abstract, MeSH descriptor terms, and publication date. All extracted data was loaded into a local relational database.

Step 4: MeSH Category Selection

Each article is assigned as many MeSH terms as appropriate to cover the topics of the article. MeSH descriptor terms are organized in 16 major categories, (e.g., category L for *Information Science* and category N for *Health Care*) [6]. Each category is further divided into subcategories, and within each subcategory, descriptors are listed in a hierarchical structure. Each descriptor has a number that indicates its tree location, e.g., computing methodologies [L01.224] is in the Information Science [L01] subcategory of the L category. We obtained the tree location numbers for each MeSH term from MeSH RDF Linked Data [10].

In this paper, we restricted our analysis on the topic trends to focus on MeSH terms in the following three categories: *Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]*, *Information Science [L]*, and *Health Care [N]*, among which, *L* category was the main focus. We chose these three categories to investigate MI research trends regarding technology and healthcare in general. Although other MeSH categories (such as diseases, humanities, geographicals) are also valuable to describe different aspects and characteristics of MI publications, they were not included in this analysis.

Step 5: Topic Trend Analysis

Detection of Core Topics: We consider a MeSH term as a core topic if it was frequently assigned to MI articles. We count the total number of publications for each individual MeSH term over the past 24 years without aggregating of the counts from its children concepts in the MeSH hierarchy. We sort the MeSH terms by the number of publications in each selected MeSH category. The top ranked MeSH terms were considered as core topics.

Detection of Trending New Topics: Novel topics were defined as those MeSH terms that debut in recent years with increasing interests by the community reflected by an increasing number of indexed articles. The determination of a MeSH term as a trending topic is done through a combined consideration of three factors: debut year, rate of increase, and number of indexed articles before 2016. The debut year is defined as the year of entry of the term into the MeSH terminology. We only included those MeSH terms debuted in the recent 10 years. The rate of increase of a MeSH terms is calculated using the formula: $a/(b+1)$, where a is the number of articles indexed by the term and published in the recent five years (2011-2015), b is the number of indexed articles published in previous five years (2006-2010), and the number “1” is a parameter to handle the situation when some MeSH terms have zero indexed articles in years (2006 – 2010). After we obtained the debut year, rate of increase, and the total number of indexed articles for each MeSH term, we were able to assign three kinds of rankings to the MeSH terms based on how recent they debut, how large the rate of increase is, and how large the total number of indexed articles is. We then obtained a combined ranking number by sum of the three ranking scores. Thereafter, we ranked the MeSH terms with the combined ranking scores and manually picked those terms that were top ranked as novel topics.

Results

Journal Analysis

After combining the journals that were renamed during their lifespan, we selected 36 journals as core MI journals (see

Figure 2). Of which, twelve journals have changed their names: “J Biomed Inform”, “Int J Med Inform”, “Inform Health Soc Care”, “Health Manag Technol”, “Comput Inform Nurs”, “AMIA Annu Symp Proc”, “IEEE J Biomed Health Inform”, “J Innov Health Inform”, “J Healthc Inf Manag”, “Top Health Inf Manage”, “J (Inst Health Rec Inf Manag)”,

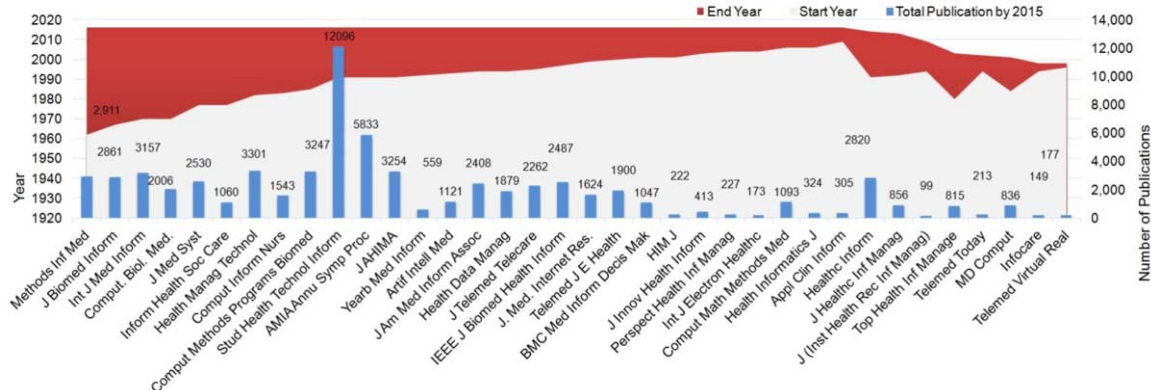


Figure 2 – Core Medical Informatics Journals with Start and End Years and Total Number of Publications.

and “Telemed J E Health”. Figure 2 shows the start and end year of each journal, as well as the total number of publications as of 12/31/2015 indexed in MEDLINE/PubMed. The average number of articles per journal is 1,834 [range: 99-12096]. The average age of the journals is 23.7 years [range: 3-55]. In Figure 2, an end year of “2016” indicates that the journal is currently active. For example, “J Biomed Inform” started publishing in 1967 under the original name “Comput. Biomed. Res.”. It was renamed using the current title in 2001 and is still active in 2016. One fifth (8 out of 36) of the journals ceased publishing or indexing to MEDLINE during our study period (1992-2015). For example, a journal titled “MD Comput” is still active but the articles are no longer indexed in MEDLINE.

Publication trends

60,862 articles were retrieved from selected MI journals in 1992-2015 (see Figure 3). The increase in the volume of MI publications between 1996 and 2006 almost follows the exponential growth trend. However, a slight decrease was observed between 2006 and 2011. Between 2011 and 2015, the growth of the publication went back to its previous exponential trajectory.

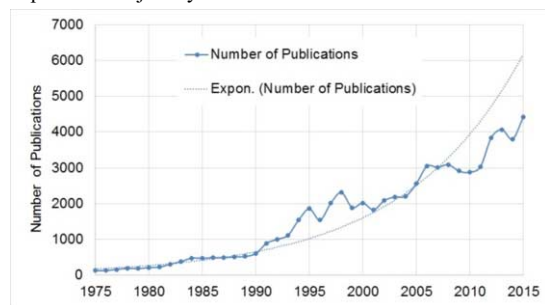


Figure 3– Publication Trends of MI Articles by Year

MeSH Analysis

We extracted a total 8,652 MeSH terms from retrieved articles, with an average of 8.3 MeSH terms per article [range: 1-32]. The MeSH terms were distributed in all 16 categories

of MeSH descriptor hierarchy, except the category of Publication Characteristics [V]. Restricting the MeSH terms to the three categories resulted in 2,929 unique MeSH terms, among which, 1,295 were in E, 378 were in L, and 1,260 were in N, while some terms overlapping categories (see Table 1). Within category L, only 68 MeSH terms belong to the branches of “Medical Informatics”, “Public Health Informatics”, and “Nursing Informatics”. Among all of the Information Science MeSH terms, we show the top 50 MeSH terms in Figure 4 sorted by the total number of indexed articles, the top 5 being: computerized medical records systems, algorithms, software, internet, and user-computer interface. The growth along every 6 years is also shown in Figure 4. Figure 5 shows trending new topics that debuted in the last decade with increasing research interests in the MI community, including: social media, mobile application, health information exchange, and cloud computing. Besides the L category, we also detected a list of core and trending new topics in the E and N categories (see Table 1).

Discussion

We identified 36 core informatics journals with over 60,000 MI articles published in the last two decades. The dissemination of MI research is currently occurring at an exponential pace. In addition, new MI journals are becoming increasingly more specialized. For example, some of the most recent MI journals include: “Telemed J E Health” (2000) focusing on publications around provision of telemedicine or telehealth; “Inform Primary Care” (2003) focusing specifically on primary care settings; and “Appl Clin Inform” (2009) providing a venue for publications focused on applied MI projects and studies.

Although we used a different set of MI articles compared to DeShazo’s work, the publication trends look similar in terms of exponential growth of publication numbers over the past 40 years. One caveat still remains – no growth was observed between 2006 and 2011, compared to the previous and following periods. One possibility is the effect of the financial crisis around 2007- 2008 and the great recession in world market during the late 2000s and early 2010s.

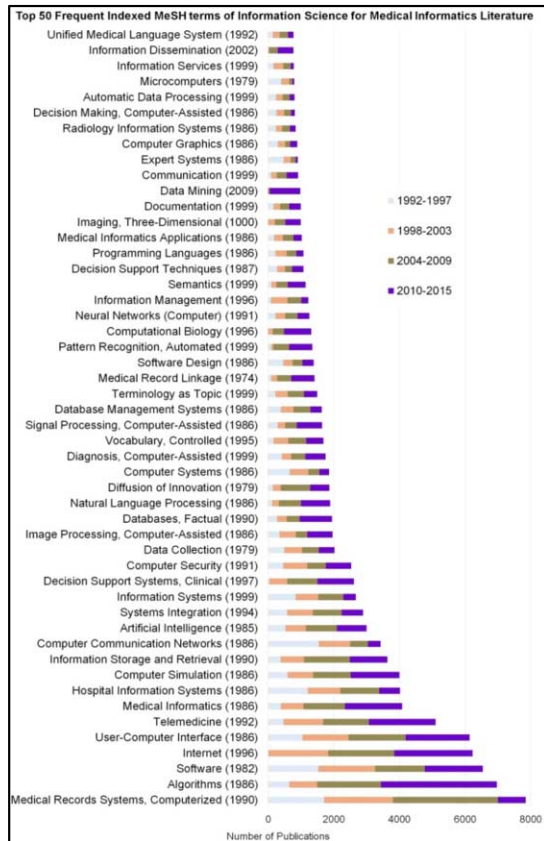


Figure 4– Core Topics in Information Science Category

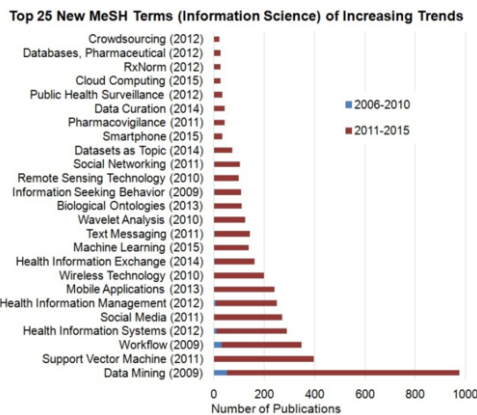


Figure 5– Novel Topics in the Information Science Category

Regardless of the stagnation period during 2006-2011, two booms in the number of publications were observed right before and after that period, 2004-2005 and 2011-2013 respectively. We suspect that these two waves may have partially been stimulated by the U.S. national health information technology (HIT) regulatory and incentive programs and similar efforts made by other countries.

In 2004, President Bush outlined a plan to ensure that most Americans have electronic health records (EHRs) within the next 10 years. It was stated in a 2005 report by the U.S. Department of Health and Human Services that "HIT is a high

Table 1– Core and New Topics for categories E and N

Cat.	E	N
Core	Electronic Health Records; Models, Theoretical; Models, Statistical	Reproducibility of Results; Surveys and Questionnaires; Delivery of Health Care; Confidentiality; Efficiency, Organizational; Attitude of Health Personnel; Pilot Projects; Models, Statistical; Organizational Case Studies; Remote Consultation; Medical Records; Quality Assurance, Health Care; Organizational Innovation; Practice Guidelines as Topic; Evaluation Studies as Topic; Patient Satisfaction; Quality of Health Care; Forms and Records Control; Patient Education as Topic
New	Health Records, Personal (2008); Precision Medicine (2008); Early Detection of Cancer (2009); Actigraphy (2010); Genome-Wide Association Study (2009); Accelerometry (2012); Brain-Computer Interfaces (2009); Clinical Alarms (2009); Health Smart Cards (2015); Drug Repositioning (2010); Clinical Decision-Making (2012); Epidemiological Monitoring (2012);	Meaningful Use (2012); Patient Safety (2011); Quality Improvement (2010); American Recovery and Reinvestment Act (2009); Clinical Coding (2010); Radio Frequency Identification Device (2009); Health Level Seven (2009); Accountable Care Organizations (2011); Patient Protection and Affordable Care Act (2010); Self Report (2010); Patient Preference (2009); Data Accuracy (2015); Patient Handoff (2012); Health Literacy (2009); Knowledge Management (2010); Comparative Effectiveness Research (2009); Medication Reconciliation (2010); Epidemics (2010); Checklist (2009); Patient Discharge Summaries (2013); Biometric Identification (2009); Tertiary Care Centers (2012); Health Communication (2010); Patient Outcome Assessment (2013); Medication Adherence (2008)

Cat. = Category, E = Analytical, Diagnostic and Therapeutic Techniques and Equipment [E], N = Health Care, Core = core medical informatics topics, New = New trending topics

priority for American healthcare system and the U.S. economy. IT is a pivotal part of transforming our healthcare system." [11] The 2011 and recent increase was perhaps stimulated by the Health Information Technology for Economic and Clinical Health (HITECH) Act, enacted as part of the American Recovery and Reinvestment Act (ARRA) of 2009, which aimed to promote the adoption and meaningful use of HIT. As a result, the initiatives have increased the adoption of EHRs across U.S. hospitals from about 30% in 2011 to almost 85% in 2015 [12]. The increased adoption in turn produced many more studies and projects, which has resulted in not only more publications but also new topics, such as meaningful use, quality improvement and ARRA, as demonstrated in Table 1.

For the MeSH terms analysis, focusing on a broader set of MeSH terms than DeShazo's helped us to identify many core topics outside of MI MeSH terms. For example, in the L category, we found a large number of publications in topics, such as algorithms, user-computer interface, natural language processing, controlled vocabulary, internet, and computer simulation, which are beyond MI subcategory. By including the general health care (N) and techniques (E) categories, our findings reveal the continuously growing core topics and newly appeared topics that medical informaticians have been dealing with in collaboration with professionals in healthcare and other technological fields.

Some core topics experienced increasing interests of the MI community in recent years (e.g., algorithms, data mining, database, pattern recognition, medical records linkage). The number of publications in several popular MeSH topics remains stable over the years (e.g., software, terminology and vocabulary, artificial intelligence). This indicates persistent interest in those topics. On the other hand, there are several

topics that show a decreasing number of publications in recent years (e.g., computer or information systems, microcomputers, and computer communication networks). This might indicate that more specific MeSH terms (along with newly added ones) were used to index articles rather than using broader terms, such as computer or information systems. Alternatively, this might indicate outdated terms with only infrequent use today (e.g., microcomputers referring to “small computers using microprocessor chips”) [13]. Overall, comparing the core topics extracted from the MI articles in the *L* and *E* categories, the majority of the terms were mentioned as part of the core content defined by Gardner *et al.* [8]

Our approach also detected new trending topics that have appeared only in the recent years. The most frequent new topic was the data mining, introduced in 2009 with almost 1,000 publications before the end of 2015. The second most frequent new topic was machine learning (including support vector machines). With the volume of health related data information growing exponentially, it is not surprising that new tools, such as data mining or machine learning are increasingly applied to process these data. Other literature also supports our findings on the increasing application of machine learning and data mining in healthcare [14–16]. Other frequent new topics in information science include concepts related to new technologies (e.g., wireless, smartphone, sensor, and mobile), as one might expect. In agreement with other literature [17,18], we found an increasing research focus on social media and related knowledge exploration techniques (e.g., crowdsourcing and pharmacovigilance) in healthcare. In terms of *Analytic, Diagnostic and Therapeutic Techniques*, electronic/personal health records are important tools to support analytics and statistical modeling, both of which show large increased interests to the MI community. In addition, recent advances in genetic testing and other molecular or cellular analysis gave birth to new topics such as precision medicine and genome-wide association study. In terms of topics in *Health Care*, two groups of trending topics were identified: (1) patient-centered care, which is indicated by the topics such as patient participation, patient preference, patient satisfaction, patient handoff, and health literacy; (2) medical safety, which is reflected by the trending terms like patient safety, medication reconciliation, medication adherence, and clinical alarms.

This study has several limitations: First, we collected articles from a set of selected MI journals. We could have missed MI articles published in journals in other fields, such as medicine, nursing, public health, library science, computer science, and engineering. However, articles published in a non-MI journal may not truly be MI, although it was assigned a MeSH term in the MI hierarchy. For example, “databases, fractal” is a MeSH term in the MI hierarchy and it can be assigned to many non-MI articles that used databases in their studies. As our study includes core MI journals, we believe that our findings capture the mainstream topics in the MI field. Second, using MeSH as part of our search and analysis strategy can be problematic. The MeSH terms may not totally reflect diverse topics in the domain of MI. The MeSH indexes can be subjective, although they were assigned by trained subject matter experts. Also, the addition of new MeSH terms often belatedly follows the actual research trend. Although these new MeSH terms may be assigned to literature retrospectively, our trend analysis based on MeSH may be chronologically behind the real trend. Third, our study does not include analyses of journal's impact factors and other bibliometrics, such as citations. Lastly, we focused on the MeSH term analysis over the last 24 years; the patterns of the trending and emerging topics detected in this study could be easily outdated as time passes.

Conclusions

We used MeSH terms to describe ongoing and emerging trends in MI publications. Our results indicate that over the past two decades, the body of MI literature has grown at an almost exponential pace. Our analysis indicates that while some popular topics in MI publications remain stable throughout the years (e.g., software, artificial intelligence), new topics also emerged in the past few years (e.g. data mining, social media). Our findings can be used to examine the existing MI educational programs (in terms of content coverage) and suggest new research avenues (e.g., emerging topics such as crowdsourcing).

Acknowledgements

This study was supported by the Agency for Healthcare Research and Quality (R01HS022728 and R01HS024264).

References

- [1] <https://www.nlm.nih.gov/hsrinfo/informatics.html>
- [2] M.F. Collen. Origins of medical informatics. *West J Med* 145(1986),778–85.
- [3] <https://www.amia.org/clinical-informatics-board-review-course/history>
- [4] https://www.amia.org/education/programs-and-courses?field_state_value=All&term_node_tid_depth=All
- [5] J.P. DeShazo, D.L. Lavallie, F.M. Wolf, Publication trends in the medical informatics literature: 20 years of “Medical Informatics” in MeSH, *BMC Med Inform Decis Mak* 9(2009),7.
- [6] <https://www.nlm.nih.gov/bsd/disted/meshutorial/me>.
- [7] A.R. Aronson, J.G. Mork, C.W. Gay, S.M. Humphrey, W.J. Rogers, The NLM indexing initiative's medical text indexer, *Stud Health Technol Inform* 107(2004),268–72.
- [8] R.M. Gardner, J.M. Overhage, E.B. Steen, B.S. Munger, J.H. Holmes, J.J. Williamson, et al. Core Content for the Subspecialty of Clinical Informatics. *J Am Med Informatics Assoc* 16(2009):153–7.
- [9] https://www.nlm.nih.gov/bsd/serfile_addedinfo.html.
- [10] <https://id.nlm.nih.gov/mesh/>.
- [11] The Lewin Group. Health information technolog leadership panel. 2005.
- [12] <https://dashboard.healthit.gov/evaluations/data-br>.
- [13] <https://www.ncbi.nlm.nih.gov/mesh/?term=microcompu>.
- [14] I. Kadi, A. Idri, J.L. Fernandez-Aleman, Knowledge discovery in cardiology: A systematic literature review, *Int J Med Inform* 97(2017) 12–32.
- [15] M. Herland, T.M. Khoshgoftaar, R. Wald, A review of data mining using big data in health informatics. *J Big Data* 1(2014),2.
- [16] D.A. Clifton, K.E. Niehaus, P. Charlton, G.W. Colopy, Health Informatics via Machine Learning for the Clinical Management of Patients, *Yearb Med Inform* 10(2015),38–43.
- [17] M.P. Hamm, A. Chisholm, J. Shulhan, A. Milne, S.D. Scott, T.P. Klassen, et al, Social media use by health care professionals and trainees: A scoping review, *Acad Med* 88 (2013) 1376–83.
- [18] L. Sinnenberg, A.M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, R.M. Merchant, Twitter as a Tool for Health Research: A Systematic Review, *Am J Public Heal* 107(2017), e1–e8.

Address for correspondence

Liqin Wang: (857) 282-4089, lwang@bwh.harvard.edu
399 Revolution Drive, Suite 1315
Somerville, MA 02145

Leveraging Event Reporting Through Knowledge Support: A Knowledge-Based Approach to Promoting Patient Fall Prevention

Bin Yao ^a, Hong Kang ^a, Qi Miao ^b, Sicheng Zhou ^a, Chen Liang ^c, Yang Gong ^a

^a School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Texas, USA

^b School of Public Health, The University of Texas Health Science Center at Houston, Texas, USA

^c Louisiana Tech University, Ruston, Louisiana, USA

Abstract

Patient falls are a common safety event type that impairs the healthcare quality. Strategies including solution tools and reporting systems for preventing patient falls have been developed and implemented in the U.S. However, the current strategies do not include timely knowledge support, which is in great need in bridging the gap between reporting and learning. In this study, we constructed a knowledge base of fall events by combining expert-reviewed fall prevention solutions and then integrating them into a reporting system. The knowledge base enables timely and tailored knowledge support and thus will serve as a prevailing fall prevention tool. This effort holds promise in making knowledge acquisition and management a routine process for enhancing the reporting and understanding of patient safety events.

Keywords:

Patient Safety; Risk Management; Accidental Falls

Introduction

Patient safety events are the most concerning issues in the improvement of healthcare quality. Patient falls, the most common safety events resulting in adverse patient outcomes and imposing significant costs, become a great burden to society. Patient safety organizations (PSOs) have listed patient falls as one of the top patient safety events [1]; 92% of in-hospital falls are preventable [2]. Efforts such as prevention and assessment tools as well as reporting systems have been developed to reduce the recurrence of patient falls in the U.S.

Prevention and assessment tools are designed to help healthcare providers assess fall risk and reduce the occurrence of fall events [3; 4]. They also provide protocols for fall prevention in terms of leadership, evaluation of fall risks (vital status, medication, environments, etc.), and patient education. However, fall prevention remains challenging due to the diversity of risk factors and localized vulnerabilities in patient risk management, which calls for tailored interventions regarding risk factors per patient [3].

Learning from reported events is critical for identifying vulnerabilities in patient safety management. Reporting systems have been implemented for collecting patient safety event data and conducting root cause analyses (RCA). Mandatory and voluntary reporting systems complementarily serve different levels and different purposes in patient safety events management [5]. To standardize reporting formats, AHRQ Common Formats (CFs) were developed and then became widely accepted and commonly used for reporting and analyzing patient safety events including falls in voluntary reporting systems [6]. The CFs fall-reporting form includes 13 structured questions which cover most contributing factors of

a fall event reporting and RCA, such as circumstances, outcome, fall risk assessment, preventions, medication, and assistant devices [7]. The reporting systems are intended to facilitate risk identification, event data collection, and RCA [8]. The ultimate aim is to advance strategies of preventing and mitigating patient safety events.

However, the success of the tools and reporting systems on patient safety has yet to be seen, since healthcare providers fail to receive timely feedback and customized knowledge support [9; 10]. The event reporting systems merely serve as data repositories if reporters can learn nothing from them [11]. Delayed and non-customized feedback may lead to a lack of incentives for providers to continuously and voluntarily report events, which in return results in the underreporting and low-quality reports [12]. This also leads to insufficient knowledge support for providers towards effective fall prevention. Consequently, a knowledge base is expected to provide the foundation for knowledge-based interventions, if one could be developed and integrated into the routine workflow of patient risk management [13]. In such a knowledge base, the solutions for patient fall should be included, and their logical connections to the specific cases should be well established to support learning and clinical decision making. Within the patient fall management circle, event reporting, retrospective analysis, and prospective analysis are chained to support each other [9]. To support the information flow from event reporting to retrospective analysis, reporting data-driven feedback is necessary. Therefore, a refined reporting system with a knowledge base that supports effective feedback has the potential to bridge the gap between the reporting process and knowledge support.

To develop the knowledge base, we identified and synthesized entry-based fall solutions from multiple resources, and established connections between the solutions and the CFs fall-reporting form through expert review. As a result, a fall prevention knowledge base including the reports, solutions, and the connections between the two datasets was established as a resource for timely tailored knowledge support in a patient fall reporting system.

Methods

We developed a framework for the knowledge base based on the CFs fall-reporting form. Fall prevention solutions gleaned from multiple resources were filtered, combined and adjusted to fit into the framework. Each solution is connected to a targeted answer option in the reporting form. Solutions and their connections to the framework were revised by expert review (Figure 1).

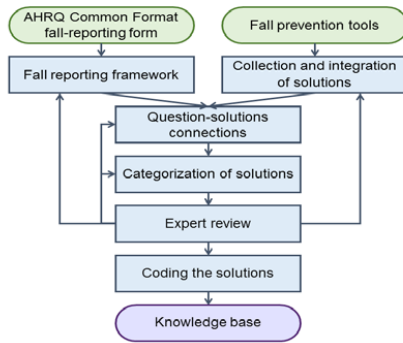


Figure 1 – Workflow of establishing the patient fall solutions knowledge base.

Establish a question-answer-solution framework for fall-reporting

To establish a network that links questions, answers, and solutions, we developed a framework of hierarchical questions for describing patient fall events based on AHRQ CFs reporting. These questions were adjusted from AHRQ CFs Hospital Version 1.2, released in April 2012 [7].

Collect solutions from multiple resources

We collected solutions from multiple tools for patient fall prevention, which were developed and distributed by Authentic Patient Safety Organizations. Two domain experts combined identical or similar solutions. Meanwhile, we searched Medline using “solutions” or “interventions” combined with “patient falls” as keywords, limited in abstracts and titles to identify additional resources containing information about patient fall solutions or interventions. The search, conducted in June 2016, resulted in 61 hits, which were integrated into the solution dataset. The following details the resources we used.

Pennsylvania Patient Safety Authority (PA PSA)

PA PSA [14] provides patient safety tools including solutions for fall prevention. The tools for falls prevention include prevention program tools and educational tools, which can serve as knowledge support for risk factors measurement and post-fall investigation. In particular, they include risk assessment tools for radiology and ambulatory surgical facilities. Pennsylvania Patient Safety Reporting System (PA-PSRS) Falls Event Type Decision Tree is provided for the determination of the fall event types in the PA-PSRS.

Joint Commission Center for Transforming Healthcare Targeted Solutions Tool (TST)

The Preventing Falls TST [15] is one of the four applications of Joint Commission TST aiming to improve the safety and quality of healthcare. It provides approaches for state evaluation, RCA, targeted solutions and strategies for sustaining patient safety management, including patient fall prevention.

Slip, Trip and Fall Prevention for Healthcare Workers

Slip, Trip and Fall Prevention for Healthcare Workers [16] was developed by Centers for Disease Control and Prevention The National Institute for Occupational Safety and Health (CDC-NIOSH). This tool listed top risk factors of patient falls, including indoor and outdoor environment conditions and improper use of equipment.

The third report from the Patient Safety Observatory. Slips, trips and falls in hospital

This is a summary of patient fall risk factors and recommended solutions based on “the slips, trips and falls in hospital, the third report from the National Patient Safety Agency’s (NPSA) Patient Safety Observatory” [1].

The “How to guide for reducing harm from falls”

This tool was developed by Patient Safety First [17], a California partnership for health. It provides fall prevention solutions for both leaders and frontline staff to implement interventions to reduce patient fall rates.

AHRQ WebM&M

AHRQ WebM&M [18] provides peer-reviewed patient safety cases and expert analysis, which can serve as a resource of patient safety event solutions. WebM&M is an example of case-specific knowledge support. Due to the store-and-forward method, the expert-reviewed case analysis is far from timely knowledge support and is merely for the purpose of case study.

AHRQ Patient Falls Prevention Toolkit

AHRQ Patient Fall Prevention Toolkit [19] contains a set of strategies and tools collected by experts from multiple sites. The toolkit provides a roadmap for stakeholders to engage in patient fall prevention. It also includes a collection of practical tools for reporters such as STRATIFY Risk Assessment Tool, Morse Fall Scale, Medication Fall Risk Score and Evaluation Tools, and other tools for mental and physical status evaluation.

Match the solutions to the Common Formats

We matched the solutions to the questions and answer options of the CFs based on the solution contents. The solutions that were successfully linked to single or multiple answer options were categorized as specific solutions. The others were categorized as general solutions. We further categorized the solutions into three different groups: Direct action, which consists of operational advice to the staff; Principle, which consists of principal guidance; Patients, which consists of knowledge support for patient education.

Expert review

An expert review was performed to ensure the solutions, connections between solutions and questions, and the categories of solutions were comprehensive for frontline practitioners. We constructed a survey to demonstrate the CF framework with relevant solutions listed below. Each solution was tagged with its category and mapping rule. Five patient safety experts working with a PSO participated in the survey. They revised the contents, categories and the mapping rules of each fall solution. Panel discussions were employed to establish consensuses.

Code the solution mapping rules

The mapping rules between solutions and answer options in CFs were coded into our knowledge base. Boolean logic connectors “AND”, “OR” and “NOT” were applied to form the preconditions for the activation of any solution in the feedback. In detail, the “AND” stands for a combination of two or multiple options need to be checked in the reporting to show the certain solutions. The “OR” requires either of the alternatives. The “NOT” means the solutions would be given in the feedback only if the certain options were not checked.

Develop a knowledge-based web server

A relational database was implemented on our server to accommodate the knowledge base containing reports,

solutions, and connections between reports and solutions. A web-based interface was developed in conjunction with the server and database.

Results

Develop a question-answer-solution framework

We set up a framework based on the 13 questions of CFs fall-reporting form. These structured questions are hierarchically organized with answer options for reporting and RCA (Table 1).

Table 1 – Questions from CFs fall-reporting form

No.	Question
1	Was the fall unassisted or assisted?
2	Was the fall observed?
3	Who observed the fall?
4	Did the patient sustain a physical injury as a result of the fall?
5	What type of injury was sustained?
6	Prior to the fall, what was the patient doing or trying to do?
7	Prior to the fall, was a fall risk assessment documented?
8	Was the patient determined to be at increased risk for a fall?
9	At the time of the fall, were any of the following risk factors present?
10	Which of the following were in place and being used to prevent falls for this patient?
11	At the time of the fall, was the patient on medication known to increase the risk of fall?
12	Was the medication considered to have contributed to the fall?
13	Did restraints, bedrails, or other physical device contribute to the fall (includes tripping over device electrical power cords)?

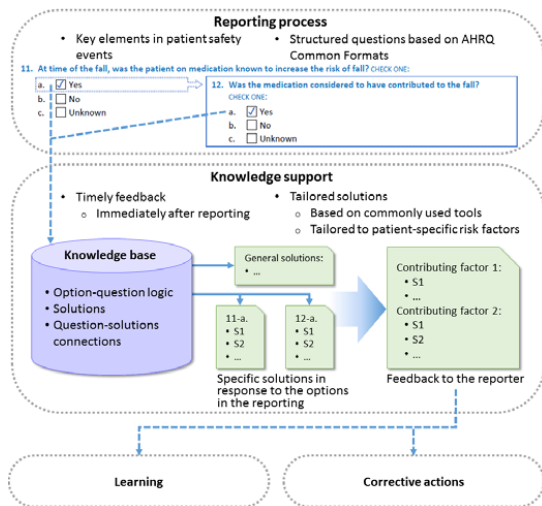


Figure 2 – The framework of the knowledge support strategy based on CFs fall reporting form.

These questions require documentation of the outcomes after fall events (Question 4-5). They further collect the risk factors (Question 6-13). The focuses of the questions can be described by 6 topics as shown in Table 1: Staff involved in the fall event (Question 1-3), outcomes of the fall (Question 4-5), circumstances of the fall (Question 6), fall risk assessment (Question 7-9, 13), fall prevention approaches (Question 10) and medication (Question 11-12). Common risk factors were included in the answer options such as medication, environment, equipment, and patient status. Logical links between options and questions were embedded to the framework. For instance, question 12 would be required

to answer if the reporter checked “Yes” for question 11 (Figure 2).

Collect and categorize fall solutions

The current tools have both overlapping and complementary contents. In general, the solutions cover the following aspects in fall prevention: assistive devices (including alarms, call lights, etc.), environment and equipment, fall event reporting, use of fall risk assessment tools, individual patient fall risks, medications, patient and family education, and rounding. As shown in Table 1, identified solutions were categorized into “general solutions” and “specific solutions” according to whether the solutions can be matched to certain answers of certain questions in the AHRQ CFs fall-reporting form. Twenty general solutions and 102 specific solutions were generated. The general solutions were categorized in 8 topics. The specific solutions were categorized in 6 topics according to the questions of CFs fall-reporting form (Table 2). In the topics of specific solutions, the count for Outcome of the fall is 0 because all of the current solutions that we collected were focused on the occurrence and recurrence of fall rather than the treatment of the outcomes which is a relatively independent process. The general solutions will be shown after all reports. Specific solutions will appear only when the corresponding answer options are chosen during the reporting. We also linked the commonly used risk factor assessment tools including Morse Scale, STRATIFY Scale and Medication Fall Risk Scale to the associated solutions, so that the reporters will be able to conduct assessments using those tools after reporting.

Table 2 – Solution categories in alphabetical order.

Contributing factors	General	Specific
Alarms	1	14
Assistive Devices	2	14
Call Lights	0	3
Environment and Equipment	7	14
Fall Event Report	1	2
Fall Risk Assessment Tool	2	8
Individual Patient Fall Risks	2	13
Medications	1	12
Patient and Family Education	2	4
Rounding	2	0
Showering/Toileting	0	10
Unassisted Ambulation	0	8
Sum	20	102

Assess recommended fall solutions

Experts’ comments from the survey were further categorized as follows:

Description improvement. Improving the description of solutions to eliminate ambiguities.

Detail specification. Adding specific terms to the solution description to provide practicable details.

New solutions. Supplementing with new solutions that were not included in the survey.

Generalization of target populations or application scopes. Expanding the target populations. For example, patient and family were included in the targets of “standardized education regarding the use of call lights” which targeted to the staff only before the expert review. The application scopes were also extended. For example, fall events without patient injuries were included in the reporting.

Workflow complementation. Implementing protocols to improve the workflow. For example, in the “Patient and family education” part, recommendation for conducting a “post-fall investigation” was added.

Solutions were improved according to the experts’ comments. Both general solutions and specific solutions were grouped by contributing factors, as shown in Table 2. By connecting the solutions with the answer options of questions in CFs fall-reporting form, a knowledge base for fall events was established. The knowledge base contains prevention solutions, connections between solutions and CFs questions, and the logical links of CFs questions (Figure 2).

Discussion

Making the reporting easy and meaningful

Though various reporting systems are available, healthcare providers face challenges using these systems. Extra training is required for providers to use the reporting systems. However, the effects of training may not live up to expectations, which makes the reporting process prone to mistakes [8]. Another problem in patient event reporting is repeated and under-qualified reports without useful information for RCA and corrective recommendations [8]. To solve these problems, we implemented the framework of AHRQ CFs fall-reporting form and integrated expert-reviewed patient fall solutions to develop a knowledge base, which provided an accessible and easy-to-use method for patient fall reporting. AHRQ CFs provided formats for structured information of patient safety events reports, which hold promise in unifying the reporting process and analysis. Based on the framework of AHRQ CFs, our system is expected to improve report quality and make reports more meaningful.

Learning from reporting

Research has been conducted to explore the proper use of fall event reports and to reduce the recurrence of patient falls [2; 3; 12]. However, traditional learning materials failed to provide effective and efficient interventions after reporting. Experienced reporters, such as expert physicians and senior nurses, may be familiar with the common solutions. However, continuous learning from practices is essential especially for novice caregivers [20].

Despite duplicate and incomplete data, current patient fall data from reporting provides rich information for reporters to learn from previous errors [21]. In the present study, we defined four levels of learning in patient safety events prevention as shown in Figure 3. National and local reporting systems collect and analyze patient safety events data, which may potentially contribute to learning on national and local scales. There are other resources for learning like guidelines, literature, and case note reviews [11]. Except for reporting systems, none of the methods is able to provide timely learning after events in individual or local level.

Therefore, a mechanism that integrates multiple solution resources and provides knowledge support after reporting is needed. In this study, we collected and integrated solutions from multiple resources for preventing patient falls. Based on the expert-reviewed solutions, we established a knowledge base that is available in a fall reporting system with timely feedback. By providing timely knowledge support after reporting, our research aims to bridge the gap between local reporting systems and local/individual learning levels (Figure 3), and to help healthcare settings identify vulnerabilities in patient risk management. As shown in Figure 3, the knowledge-based reporting system helps facilitate the

information flow from the reporting process to the learning process and then the process of identification and prevention of patient safety events (Figure 3).

The diversity of organizational weaknesses in patient safety management is a challenge for healthcare providers to apply prevention solutions. Therefore, it is essential for patient safety managers to identify local vulnerabilities in management and perform corrective measures. The feedback after reporting can provide not only immediate post-fall solutions to the reporters and help prevent the recurrence of similar patient fall events but also alerts patient safety managers about management system vulnerabilities.

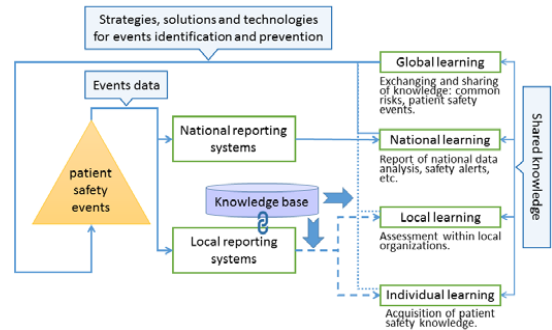


Figure 3 – Interactions between patient safety events reporting and learning in different levels. Reporting systems collect and analyze patient safety events data, provide information and knowledge for learning in different levels.

Bridging the gap in patient fall management

The three stages of the patient safety management circle (i.e., prospective analysis stage, surveillance stage, and retrospective analysis stage) overlap with each other. Correspondingly, an ideal pattern is integrating the reporting process and knowledge support into the existing workflow of patient fall management, which facilitates learning from reporting as well as the assessing and correcting processes. However, the gap between reporting and knowledge support appears to be a major barrier to establishing an integrated management mechanism for patient safety. A practical approach is to explore the potential connections between various usable tools of fall prevention and the reporting systems so that reporters will be able to gain timely feedbacks after reporting. Our approach provided a method to integrate expert-reviewed patient fall solutions into the reporting system and to offer timely knowledge support.

For other types of events, a specific knowledge base should be established within the corresponding domains to connect events and solutions. As a result, the strategy for establishing a knowledge base that was used here could be adapted accordingly. For example, pressure ulcers are an injury to the skin and underlying tissue, and are described in four stages according to severity. For each stage, solutions and mapping rules to reporting forms should be collected and evaluated separately to ensure the practicality of knowledge support. Establishing a knowledge base for medication reconciliation could also be considered. The knowledge base should consider the drug name, dosage, frequency, and route while archiving the solution entries. Moreover, physicians' admission, transfer, and/or discharge orders should also be considered while connecting medication events to solutions. Therefore, future patient safety reporting systems with timely knowledge support should be based on a group of knowledge bases in various domains.

Limitations

The solutions collected in this study may not have covered all aspects of patient falls. Additionally, there are still controversial opinions on the effectiveness and necessity of several solutions due to the lack of large-sample experiments. Further evaluation of such solutions is needed to mitigate such controversy. Although AHRQ CFs cover most key elements for event reporting and RCA, some aspects may be not included due to the diversity of healthcare settings.

The efficiency and effectiveness of our feedback mechanism also need to be further evaluated. Considering the time-consuming and multiple-step procedures of assessment and reporting, the risk factor assessment and knowledge support still call for collaborations from healthcare providers, patient safety experts, and informaticists.

Another challenge is developing and embedding a human-computer communication mechanism into the healthcare providers' workflow. A critical barrier for patient event reporting is the reluctance of providers to report because of a potentially discouraging environment and workflow interruption. In some cases, reporters were not the personnel relevant to the events who would most benefit from learning from reporting.

Future work

Future work should extend knowledge base development and management of patient falls to those of other event types. The mechanism of sharing and learning from safety events would enhance the reporting quality at all levels. In addition, maintaining and updating the knowledge base can be challenging and labor-intensive, but procedures maintain updated knowledge bases are necessary.

Conclusion

As more high-quality data is collected, stored and analyzed by reporting systems, patient falls can be an important use case for establishing an integrated and optimized workflow using informatics strategies. Our work explored the possibility of building connections between reporting and knowledge support via a knowledge base. This approach may be extensible to other fields of patient safety events, such as pressure ulcers, medication errors, health information technology (HIT) issues, surgical safety, hand hygiene, and medical device issues.

Acknowledgements

This project is supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), Agency for Healthcare Research & Quality (1R01HS022895), and University of Texas System Grants Program (#156374).

References

- [1] F. Healey and S. Scobie, The third report from the Patient Safety Observatory. Slips, trips and falls in hospital, in, National Patient Safety Agency, London, 2007.
- [2] J.M. Morse, *Preventing Patient Falls: Establishing a Fall Intervention Program*, NY: Springer, New York, 2008.
- [3] P.C. Dykes, D.L. Carroll, A. Hurley, S. Lipsitz, A. Benoit, F. Chang, S. Meltzer, R. Tsurikova, L. Zuyov, and B. Middleton, Fall Prevention in Acute Care Hospitals: A Randomized Trial, *The Journal of the American Medical Association* **304** (2010), 1912-1918.
- [4] A. Carson-Stevens, P. Hibbert, H. Williams, H.P. Evans, A. Cooper, P. Rees, A. Deakin, E. Shiels, R. Gibson, A. Butlin, and B. Carter,

- Characterising the nature of primary care patient safety incident reports in the England and Wales National Reporting and Learning System: a mixed-methods agenda-setting study for general practice, in, 2016.
- [5] L. Kohn, J. Corrigan, and M. Donaldson, *To Err Is Human: Building a Safer Health System*, in, Washington, DC, 2000.
 - [6] C.M. Clancy, Common Formats Allow Uniform Collection and Reporting of Patient Safety Data by Patient Safety Organizations, *American Journal of Medical Quality* **25** (2010), 73-75.
 - [7] Common Formats - Version 1.2, in, Agency for Healthcare Research and Quality, 2012.
 - [8] J.C. Pham, T. Girard, and P.J. Pronovost, What to do with healthcare Incident Reporting Systems, *Journal of Public Health Research* **2** (2013), 154-159.
 - [9] L. Hua and Y. Gong, Information gaps in reporting patient falls: the challenges and technical solutions, *Stud Health Technol Inform* **194** (2013), 113-118.
 - [10] C. Stavropoulou, C. Doherty, and P. Tosey, How Effective Are Incident-Reporting Systems for Improving Patient Safety? A Systematic Literature Review, *The Milbank Quarterly* **93** (2015), 826-866.
 - [11] I. Larizgoitia, M.-C. Bouesseau, and E. Kelley, WHO efforts to promote reporting of adverse events and global learning, *Journal of Public Health Research* **2** (2013), 168-174.
 - [12] S. Hempel, Hospital Fall Prevention: A Systematic Review of Implementation, Components, Adherence, and Effectiveness, *Journal of the American Geriatrics Society* **61** (2013), 483-494.
 - [13] S.D. Wilson, Nurses' Perceptions of Implementing Fall Prevention Interventions to Mitigate Patient-Specific Fall Risk Factors, *Western Journal of Nursing Research* **38** (2016), 1012-1034.
 - [14] Patient Safety Tools, in, Pennsylvania Patient Safety Authority, 2013.
 - [15] J. O'Donovan and A. Bersin, Controlling Ebola through mHealth strategies, *Lancet Glob Health* **3** (2015), e22.
 - [16] Slip, Trip, and Fall Prevention for Healthcare Workers, in, Centers for Disease Control, 2011.
 - [17] P.S. First, The How to guide for reducing harm from falls, in, National Patient Safety Agency, London, 2009.
 - [18] S. Agarwal, H.B. Perry, L.A. Long, and A.B. Labrique, Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review, *Trop Med Int Health* **20** (2015), 1003-1014.
 - [19] E.D. Spikol, C.E. Laverriere, M. Robnett, G. Carter, E. Wolfe, and E. Glasgow, Zebrafish Models of Prader-Willi Syndrome: Fast Track to Pharmacotherapeutics, *Diseases* **4** (2016).
 - [20] N. Rafter, A. Hickey, S. Condell, R. Conroy, P. O'Connor, D. Vaughan, and D. Williams, Adverse events in healthcare: learning from mistakes, *The Quarterly journal of medicine* **108** (2015), 273-277.
 - [21] J.R. Clarke, The use of collaboration to implement evidence-based safe practices, *Journal of Public Health Research* **2** (2013), 150-153.

Address for correspondence

Yang Gong, MD, PhD
7000 Fannin St Suite 165,
Houston, TX 77030, USA,
Yang.Gong@uth.tmc.edu,
Phone: +1 713 500 3547,
Fax: +1 713 500 3907.

Summarizing an Ontology: A “Big Knowledge” Coverage Approach

Ling Zheng, Yehoshua Perl, Gai Elhanan, Christopher Ochs, James Geller, Michael Halper

College of Computing, New Jersey Institute of Technology, Newark, NJ 07102-1982, USA

Abstract

Maintenance and use of a large ontology, consisting of thousands of knowledge assertions, are hampered by its scope and complexity. It is important to provide tools for summarization of ontology content in order to facilitate user “big picture” comprehension. We present a parameterized methodology for the semi-automatic summarization of major topics in an ontology, based on a compact summary of the ontology, called an “aggregate partial-area taxonomy”, followed by manual enhancement. An experiment is presented to test the effectiveness of such summarization measured by coverage of a given list of major topics of the corresponding application domain. SNOMED CT’s Specimen hierarchy is the test-bed. A domain-expert provided a list of topics that serves as a gold standard. The enhanced results show that the aggregate taxonomy covers most of the domain’s main topics.

Keywords:

Big Knowledge; Topic Coverage; Ontology Summarization

Introduction

The Big Data to Knowledge (BD2K) initiative is expected to produce many knowledge items that can be expressed as assertions or as rules. However, orientation into large knowledge bases is a challenge by itself, the “Big Knowledge” challenge. Without some high-level mental representation about the kinds of content in a large knowledge base, effective use of the knowledge may be limited [1]. When an ontology surpasses thousands of assertions, even its curators are confronted with the problem of seeing the “big picture” of its content. The work of curators, in charge of developing and maintaining the ontology, relies on its comprehension. A topic of an ontology represented by a concept c is considered a major topic if c has a large number of descendants. The effectiveness of tools for summarizing and supporting users’ comprehension of a large ontology can be measured by “Big Knowledge” coverage of a given list of major topics related to the corresponding domain. “Big Knowledge” coverage is defined as the percentage of the number of major topics in an ontology out of a given list, rather than by the percentage of the number of concepts out of a given list of concepts [2].

We summarize the “big picture” of an ontology by automatically deriving concept groups that represent major topics in a specific domain. The size of a topic, represented by a concept c in an ontology, is the number of c ’s descendants. The knowledge representation importance of topic c (which is different from its clinical importance) can be approximated by its size. The derivation of the topic-defining concepts of an ontology is based on its *aggregate partial-area taxonomy* (“*aggregate taxonomy*,” for short) [3]. The aggregate taxonomy is a significantly smaller network that summarizes

an ontology’s concepts, providing a simplified view of the ontology. Aggregate taxonomies are based on *partial-area taxonomies* (“*taxonomies*” for short) [4], a compact summarization abstraction network developed, for example SNOMED CT [5].

In this paper, we demonstrate why *aggregate taxonomies* are better at summarizing an ontology than *taxonomies*, when measuring the effectiveness of the former by comparing with a gold standard list of major topics selected by a domain expert. We used SNOMED CT’s *Specimen* hierarchy as a test-bed.

Background

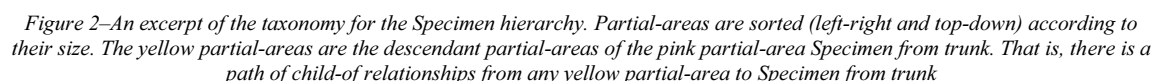
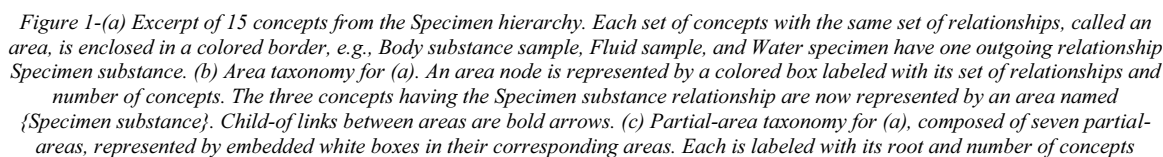
Biomedical ontologies provide terminological support for Electronic Health Records (EHRs) [6], decision-support systems, natural language processing, data integration [7], etc. SNOMED CT is a comprehensive ontology organized into 19 hierarchies, such as *Clinical finding*, *Procedure* and *Specimen*. Its 317,057 active concepts (July 2015 release) are linked by *IS-A* relationships and more than 1.5 million attribute relationships (“*relationships*,” for short). There are 1,620 concepts in the *Specimen* hierarchy.

We have developed *partial-area taxonomies* [4] to provide a compact view of the seven SNOMED CT hierarchies, including *Specimen*, that have outgoing relationships. We illustrate the process of the partial-area taxonomy derivation for an excerpt of the *Specimen* hierarchy (Figure 1(a)).

We start with the definition of the *area taxonomy*. We define an *area* as a set containing all concepts having the same set of outgoing relationships, without considering the ranges of those relationships. Areas are disjoint (i.e., each concept appears in only one area). Areas are named by their sets of relationships. Areas are represented by nodes of the area taxonomy. Figure 1(b) shows the four areas derived from the concepts in Figure 1(a). Areas are color-coded according to the cardinalities of their sets of relationships (i.e., all areas with the same number of relationships have the same color). For example, both areas {*Specimen source topography*} and {*Specimen substance*} in Figure 1(b) have one relationship, so both are colored green.

We define a *root* of an area as a concept having no parents in the area. Areas may have multiple roots. Area nodes are connected hierarchically, using *child-of* links, to form the area taxonomy. Those links are based on the configurations of area roots in the underlying ontology. More specifically, an area A is *child-of* an area B if a root in A has a parent in B . Figure 1(b) has four *child-of*s derived in this manner.

A *partial-area taxonomy* (“*taxonomy*” for short) is derived from the area taxonomy as follows. We group a root and all its intra-area descendants into a *partial-area*, which is drawn as an embedded node of the area (see Figure 1(c)). If an area has



there are relatively more concepts for a topic than it is “more important.” For example, there are 262 concepts related to digestive system specimens, but only 12 related to bone marrow specimens. We thus consider the topic “digestive system samples” as more important in SNOMED CT. We do not claim necessarily that it is more important clinically since this depends on clinical context.

Our approach for evaluating the automatically identified major topics is based on a gold standard list. A domain expert (GE) was asked to select a list of major topics for the specimen domain. (GE) is an MD with long experience in ontologies. A gold standard may also be derived from a published ontology of an authoritative organization. We did not find another ontology for specimens (e.g., in UMLS. For the sake of normalization and to simplify the eventual matching task), each chosen topic was semi-automatically mapped to a SNOMED CT concept in the *Specimen* hierarchy, utilizing UMLS synonyms. For example, the topic “Bone specimen” was mapped to the concept *Specimen from bone*.

Methods

First, we assume that concepts belonging to a given topic are all hierarchically related (i.e., they share a common ancestor concept c that represents and names the topic). That is, all the descendant concepts of a topic c belong to that topic since they are specification of c . And second, we assume that if

Specimen has 59 children (e.g., *Biopsy sample* and *Blood specimen*). However, among the 59 children, many would not be considered major topics (based on our second assumption above), since they have few descendants. For example, 13 of *Specimen*'s children do not themselves have children (e.g., *Muscle specimen*). Nine have few children and no grandchildren (e.g., *Fibroblast specimen* has one child). Of the remaining 37 children, only 13 were in the major topic list of our domain expert, while another eight on that list were not children of *Specimen* (e.g., *Stool specimen* is a grandchild of *Specimen*). Hence, a better methodology for identifying major topics is required. In this study we address two questions: (1) how well do the partial-areas in a taxonomy match the topics in the list by a domain expert?, and (2) can we modify a taxonomy to automatically capture major topics in the ontology?

Partial-Area Taxonomies for Topic Identification

As the originators of partial-area taxonomies, we note that they were not designed for major topic identification, but for structure and content summarization. Indeed, the roots of partial-areas are not necessarily intuitive topics. The root is distinguished by the introduction of a new relationship type into the ontology, which may, but is not guaranteed to, correlate with a major topic. Moreover, a partial-area may be small, and thus, may not define a broad topic. A taxonomy typically has many small partial-areas [3]. As a result, the taxonomy for a large ontology, although smaller by an order of magnitude than the ontology, can still fail to identify major topics. Metaphorically, the “forest” summary of the topics is not seen for the many small “trees” (see Figure 2).

Hence, a better solution for identifying major topics is to pick *only the large partial-areas* (with, e.g., dozens or more concepts). To illustrate these points, Figure 2 shows an excerpt of the *Specimen* taxonomy. Some concepts appear as (labels of) relatively large partial-areas. For example, *Specimen from trunk* (132), *Specimen from head and neck structure* (53), and *Specimen from digestive system* (50) from the area {*Specimen source topography*} are partial-areas with 50 or more concepts. However, all the seven large partial-areas account for only 536 *Specimen* concepts (33.1%). One may wonder about the topics of the other 66.9% of concepts.

Moving to medium-sized partial-areas with 20–49 concepts, we find eight partial-areas covering 218 (13.5%) concepts (e.g., *Blood specimen* (28) and *Soft tissue biopsy sample* (23)). Together, the large and medium partial-areas cover only 754 specimen concepts (46.5%). There are other problems with the summarization provided by the large/medium partial-areas. For example, all descendant partial-areas (yellow) of *Specimen from trunk* (pink) in Figure 2 contain refinements of this topic. They are in a separate partial-area because they have an extra relationship and appear in another area. For example, *Swab from abdomen* (13) has an additional *Specimen procedure* relationship. Overall, there are 201 partial-area descendants of *Specimen from trunk*, covering 551 concepts.

If we only focus on large and medium partial-areas, we are ignoring useful knowledge that is distributed among the many small partial-areas. Frequently, a large partial-area has many descendant small partial-areas. The concepts in these descendant partial-areas cover the same topic as the large parent/ancestor partial-area, but in more detail. Hence, they could also be summarized by the parent/ancestor partial-area.

Weighted Aggregate Partial-Area Taxonomies

We introduce an aggregation process that allows small partial-

areas to contribute to the identification of major topics. Large partial-areas are used as candidates for topics, while small ones are folded into their larger ancestor partial-area(s). Thus, the lost knowledge in small partial-areas is accounted for.

To address the inclusion of small partial-areas formally, a variation of a partial-area taxonomy based on an *adjustable minimum grouping threshold value* was introduced [3]. It is called the *aggregate partial-area taxonomy* (or just “aggregate taxonomy”). The derivation of an *aggregate taxonomy* begins with the selection of a threshold b , indicating the minimum size of a partial-area that will be included in the aggregate taxonomy. Then, using a *topological sort*, the aggregate taxonomy is generated by aggregating any partial-areas with sizes below b into their parent/ancestor partial-area(s) with sizes $\geq b$. (The root partial-area may be any size.) The nodes of this taxonomy are called *aggregate partial-areas*.

However, there is still another problem due to the structure of a *partial-area taxonomy*. We discovered that some major topics did not appear in the aggregate taxonomies at all, due to the small sizes of their partial-areas, in spite of having many small descendant partial-areas belonging to the same topic. For example, the partial-area capturing the topic *Endocrine sample* has only 10 concepts, but many more descendant concepts belong to this topic.

To overcome this difficulty, we defined an **aggregated weight** for each partial-area. This aggregated weight equals the sum of the size x of the partial-area itself and the sizes of all its descendant partial-areas smaller than x . In this way, the decision of which “small partial-areas” to eliminate from the aggregate taxonomies can now be based on the aggregated weight of a partial-area, rather than its size.

This new taxonomy is called **weighted aggregate partial-area taxonomy**. For example, the partial-area *Endocrine sample* (10) in Figure 3(a) does not appear in the aggregate taxonomy when $b > 10$. However, its aggregated weight is 26, because it has 9 descendant partial-areas with fewer than 10 concepts, summarizing 16 descendant concepts. Therefore, the partial-area *Endocrine sample* will appear in the weighted aggregate partial-area taxonomy as long as $b \leq 26$ (Figure 3(b)).

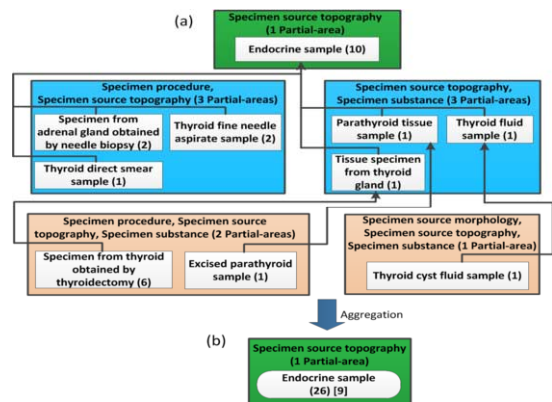


Figure 3– (a) An excerpt of 10 partial-areas. (b) Weighted aggregate partial-area for (a) with $b=7$ shown as a rounded white rectangle with its number of concepts in (), including all concepts from aggregated partial-areas, and the number of aggregated partial-areas in []

Table 1– Identification results for 21 chosen topics in weighted aggregate taxonomies with different thresholds b

Topic	Concept	Partial-area	Weight	1	5	10	15	20	25	30
Blood specimen	<i>Blood specimen</i>	<i>Blood specimen</i> (28)	43	✓	✓	✓	✓	✓	✓	✓
Body substance sample	<i>Body substance sample</i>	<i>Body substance sample</i> (63)	498	✓	✓	✓	✓	✓	✓	✓
Fluid sample	<i>Fluid sample</i>	<i>Fluid sample</i> (50)	257	✓	✓	✓	✓	✓	✓	✓
Bone marrow specimen	<i>Bone marrow specimen</i>	<i>Bone marrow specimen</i> (8)	13	✓	✓	✓	–	–	–	–
Bone specimen	<i>Specimen from bone</i>	<i>Musculoskeletal sample</i> (15)	44	–	–	–	–	–	–	–
Specimen from nervous system	<i>Specimen from nervous system</i>	<i>Specimen from nervous system</i> (12)	42	✓	✓	✓	✓	✓	✓	✓
Dermatological specimen	<i>Dermatological sample</i>	<i>Dermatological sample</i> (8)	30	✓	✓	✓	✓	✓	✓	✓
Device specimen	<i>Device specimen</i>	<i>Device specimen</i> (19)	40	✓	✓	✓	✓	✓	✓	✓
Digestive system specimen	<i>Specimen from digestive system</i>	<i>Specimen from digestive system</i> (50)	126	✓	✓	✓	✓	✓	✓	✓
Endocrine system specimen	<i>Endocrine sample</i>	<i>Endocrine sample</i> (10)	26	✓	✓	✓	✓	✓	✓	–
Genital system specimen, male	<i>Male genital sample</i>	<i>Specimen from trunk</i> (132)	489	–	–	–	–	–	–	–
Genitourinary specimen	<i>Genitourinary sample</i>	<i>Specimen from trunk</i> (132)	489	–	–	–	–	–	–	–
Hair specimen, scalp	<i>Hair specimen</i>	<i>Dermatological sample</i> (8)	30	–	–	–	–	–	–	–
Musculoskeletal specimen	<i>Musculoskeletal sample</i>	<i>Musculoskeletal sample</i> (15)	56	✓	✓	✓	✓	✓	✓	✓
Skin specimen	<i>Specimen from skin</i>	<i>Dermatological sample</i> (8)	30	–	–	–	–	–	–	–
Soft tissue specimen	<i>Soft tissue sample</i>	<i>Soft tissue sample</i> (21)	92	✓	✓	✓	✓	✓	✓	✓
Cardiovascular sample	<i>Cardiovascular sample</i>	<i>Cardiovascular sample</i> (12)	28	✓	✓	✓	✓	✓	✓	–
Specimen from eye	<i>Specimen from eye</i>	<i>Specimen from head and neck structure</i> (53)	196	–	–	–	–	–	–	–
Specimen from joint	<i>Joint sample</i>	<i>Musculoskeletal sample</i> (15)	56	–	–	–	–	–	–	–
Lesion sample	<i>Lesion sample</i>	<i>Lesion sample</i> (17)	118	✓	✓	✓	✓	✓	✓	✓
Stool specimen	<i>Stool specimen</i>	<i>Body substance sample</i> (63)	498	–	–	–	–	–	–	–
# Identified topics (C)				13	13	13	12	12	12	10

Major Topic Identification

We experiment iterating the threshold b over the range 1...30 and generate the weighted aggregate taxonomy for each b . Each such weighted aggregate taxonomy is inspected to determine its effectiveness in capturing major topics. Precision, recall, and F measure [9] were calculated for each weighted aggregate taxonomy, with the expert's topic list serving as a gold standard.

As a preliminary experiment, we tested how many of the gold standard topics appeared as partial-areas in the taxonomy (not the weighted aggregate taxonomies). We found that, out of the 21 topics chosen by the expert, 13 appear as partial-areas. This yields a recall of 0.62 (13/21) and, with 503 partial-areas in the taxonomy, very low precision of 0.03 (13/503). Note that many partial-areas are very small. In contrast, the weighted aggregate taxonomy, which eliminates the small partial-areas, is more effective. To balance recall and precision, we chose the weighted aggregate taxonomy with the b value that maximizes the F measure.

Results

If the root concept r of a partial-area appears in the weighted aggregate taxonomy of threshold b , then r is considered a topic identified by that weighted aggregate taxonomy, a corresponding checkmark “✓” is placed in Table 1. Otherwise, a dash “–” is written. For example, the topic *Bone marrow specimen* is captured by a partial-area *Bone marrow specimen* (8) with an aggregated weight 13 (Table 1). Therefore, it is identified by all weighted aggregate taxonomies with $b \leq 13$ ($b=1, 5, 10$) However, for $b>13$, *Bone marrow specimen* (8) is folded into an ancestor partial-area and disappears. No weighted aggregate taxonomy with $b>13$ identifies the topic *Bone marrow specimen*. As another example, *Bone specimen* was not identified by the weighted aggregate taxonomy with any b value as major topic (Row 5 of Table 1), since its mapped SNOMED CT concept *Specimen from bone* (Row 5, Column 2 of Table 1) is not a root of a partial-area.

At the bottom of Table 1, we show the totals of the identified

topics for the respective taxonomies. For example, for $b=5$, the total is 13. Table 2 shows each weighted aggregate taxonomy's number of partial-areas (A), recall, precision, and F. Recall is the ratio of identified topics and total topics ($R=C/S$, where $S=21$). Precision is the ratio of the identified topics and the number of partial-areas ($P=C/A$). For example, for $b=25$, the number of partial-areas is 29, the number of identified topics is 12, $R=0.57$, $P=0.41$ & $F=0.48$. Table 2 shows that $b=25$ yields the taxonomy where F is maximized. In this case, the weighted aggregate taxonomy captures 12 of the 21 topics. Figure 4 shows this weighted aggregate taxonomy with the 12 partial-areas identifying topics highlighted in yellow. The total number of concepts in these 12 aggregate partial-areas is 988, accounting for 61.0% (988/1620) of the concepts in the *Specimen* hierarchy.

Table 2–Parameterize performance of weighted aggregate taxonomies for various thresholds

$b =$	1	5	10	15	20	25	30
# Identified topics	13	13	13	12	12	12	10
# Partial-areas (A)	503	89	54	40	35	29	26
Recall ($R = C/S$)	0.62	0.62	0.62	0.57	0.57	0.57	0.48
Precision ($P = C/A$)	0.03	0.15	0.24	0.30	0.34	0.41	0.38
$F = 2 \cdot P \cdot R / (P + R)$	0.05	0.24	0.35	0.39	0.43	0.48	0.43

As an ancillary experiment, we carried out a feedback step with our domain expert (GE). When inspecting the weighted aggregate taxonomy for threshold b , one can assess whether its other partial-areas beyond those in the gold standard list are worthy of the designation “major topic”, for example, those aggregate partial-areas (Figure 4) categorizing over 25 concepts, but not in the given list. Since some topics may have been overlooked originally due to various reasons, e.g. *Specimen from head and neck structure*, a compound topic name with two body parts, and *Tissue specimen obtained by excision*, corresponding to two relationships *Specimen procedure* and *Specimen source topography*. Figure 4 was shown to (GE). He manually determined that 13 more partial-areas, highlighted in pink, warranted inclusion in the list of major specimen topics, while the other three (in white) are deemed as non-major topics. Reevaluating the experiment (with $21+13=34$ major topics), we obtained $R=0.74$

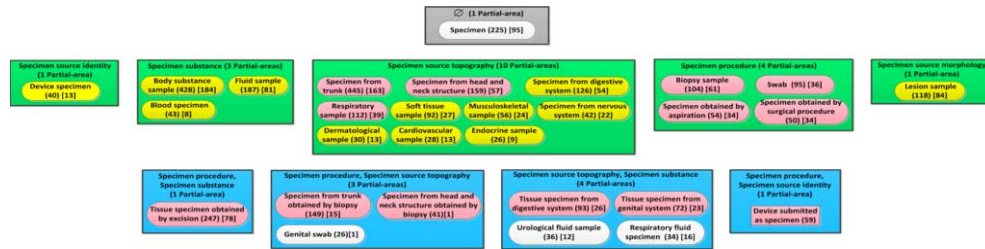


Figure 4-Weighted aggregate taxonomy for the Specimen hierarchy with $b=25$. The 12 partial-areas corresponding to the original given topics are highlighted in yellow. The 13 topics added during the enhancement step are highlighted in pink

($=25/34$), $P=0.86$ ($=25/29$) and $F=0.79$ for $b=25$. The number of concepts in these 25 aggregate partial-areas is 1,524 (94.1% of the concepts in the *Specimen* hierarchy).

Discussion

Summarizing a large ontology is a challenge as there is a lack of an objective universally accepted criteria for what constitutes a “good summarization” of an ontology. Various applications require different summaries of various granularities. Nevertheless, the management of ontologies requires “big picture” comprehension that can be enabled by compact summarization networks such as *weighted aggregate taxonomies*. Our technique is applicable to any ontology. We demonstrated our technique using SNOMED CT because of its importance in clinical applications and its large size.

We used a knowledge-oriented approach, where the importance of a topic is based on the number of concepts related to that topic in an ontology. To measure the quality of our summarizations we compared the number of identified major topics with a gold standard list of topics selected by a domain expert, who selected topics from a clinical perspective. The performance of our technique was optimized by maximizing the F symmetric measure. We modified partial-area taxonomies into weighted aggregate taxonomies to overcome difficulties in bridging the gap between the clinical perspective of the given gold standard list and the knowledge perspective of our structural methodology, which yielded better results.

A future problem is accounting for major topics that are summarized by large partial-areas (e.g., *Male genital sample* is in *Specimen from trunk* (132)). Such a topic is missed by a taxonomy since it does not appear as a root of a partial-area. It would also be more objective to obtain the gold standard list from another ontology (or other authoritative sources) rather than from one of the authors (GE). Thus, we plan to experiment with the *Disease, Disorder or Finding* hierarchy of NCI, using ICD-10 as a source of the topic list. We are also planning evaluation studies to compare our technique to related summarization methods, such as key-concept-based ontology summarization [10] and information-content-based approaches [11].

Conclusions

Summarizing knowledge bases for maintenance and “big picture” comprehension is a Big Knowledge challenge. Taxonomies were created to summarize structure and content of ontologies. We modify taxonomies to properly support “big picture” comprehension. For this, we presented the *weighted aggregate partial-area taxonomy* and

measured its performance for “Big Knowledge” coverage of a domain expert’s gold standard list of major topics for SNOMED CT’s *Specimen* hierarchy. Our methodology was parameterized in terms of a threshold b for node size to obtain optimal performance by maximizing the symmetric F measure balancing recall and precision. A manual enhancement step improved the algorithmic results.

Acknowledgments

Research reported here was supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA190779. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- [1] V.L. Patel, D.R. Kaufman, and J. Arocha, Conceptual change in the biomedical and health sciences domain, *Advances in instructional psychology* **5** (2000), 329–92.
- [2] K.W. Fung, R. Richesson, and O. Bodenreider, Coverage of rare disease names in standard terminologies and implications for patients, providers, and research, *AMIA Symp* (2014), 564–72.
- [3] C. Ochs, Y. Perl, J. Geller, and M.A. Musen, Using aggregate taxonomies to summarize SNOMED CT evolution, in: *BIBM Conference*, 2015, pp. 1008–15.
- [4] Y. Wang, M. Halper, H. Min, Y. Perl, Y. Chen, and K.A. Spackman, Structural methodologies for auditing SNOMED, *J Biomed Inform* **40** (2007), 561–81.
- [5] M.Q. Stearns, C. Price, K.A. Spackman, and A.Y. Wang, SNOMED clinical terms: overview of the development process and project status, *AMIA Symp* (2001), 662–66.
- [6] K. Giannangelo and S.H. Fenton, SNOMED CT survey: an assessment of implementation in EMR/EHR applications, *Perspect Health Inf Manag* **5** (2008), 7.
- [7] D.L. Rubin, N.H. Shah, and N.F. Noy, Biomedical ontologies: a functional perspective, *Brief Bioinform* **9** (2008), 75–90.
- [8] N. Li, E. Motta, and M. d’Aquin, Ontology summarization: an analysis and an evaluation, in: *IWEST*, 2010.
- [9] C.D. Manning, P. Raghavan, and H. Schütze, Evaluation in information retrieval, in: *Introduction to Information Retrieval*, Cambridge University Press, 2008, pp. 151–75.
- [10] S. Peroni, E. Motta, and M. d’Aquin, Identifying key concepts in an ontology, through the integration of cognitive principles with statistical and topological measures, in: *ASWC*, 2008, 242–56.
- [11] G.K. Mazandu and N. J. Mulder, Information Content-Based Gene Ontology Functional Similarity Measures: Which One to Use for a Given Biological Data Type? *PLoS One* **9** (2014), e113859.

Address for Correspondence

Ling Zheng, lz265@njit.edu, NJIT, University Heights, Newark, NJ 07102-1982 USA

KQA: A Knowledge Quality Assessment Model for Clinical Decision Support Systems

Seyedjamal Zolhavarieh^a, Dave Parry^a

^a Department of Computer Science,
Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand

Abstract

Informatics researchers have developed many methods for using computers to utilize knowledge in decision making in the form of clinical decision support systems (CDSSs). These systems can enhance human decision making in the healthcare domain. The knowledge acquisition bottleneck is one of the well-known issues in developing knowledge-based systems such as CDSS. It can be considered as a flow of knowledge from different knowledge sources to the main system. Most existing methods for extracting knowledge from knowledge resources suffer from the lack of a proper mechanism for extracting high-quality knowledge. In this paper, we propose a framework to discover high-quality knowledge by utilizing Semantic Web technologies.

Keywords:

Decision Support Systems, Clinical; Knowledge Management, Information Storage and Retrieval

Introduction

Decision making is an essential activity for clinicians in the healthcare domain. Since 1954, Clinical Decision Support Systems (CDSSs) have been developed to enhance health care systems and improve human decision-making [1]. CDSS is a particular type of decision support system [2] that guides experts in the decision-making process via electronically stored clinical knowledge [3-4]. These systems might use different approaches to assist patients by using alerts, reminders, interpretation system, etcetera.

The CDSS is built from a knowledge base (KB), inference/reasoning engine, and user/communication interaction [5]. It receives patient data and inquiry as inputs and generates a decision as an output. In this scenario, the KB plays an important role in collecting, classifying and sharing knowledge [6].

The knowledge acquisition (KA) bottleneck is one of the well-known issues in CDSS [7]. It is the process of capturing knowledge from external knowledge sources [8]. It is vital to provide an appropriate platform for interacting CDSSs and KBs. Every CDSS needs to rely on high-quality knowledge retrieved from KBs since the CDSS will not be effective if it uses out of date, limited or incomplete knowledge [9]. In addition, finding the latest accurate clinical knowledge to support decision-making is difficult. This issue is partly due to the enormous amount of research, guidelines and other knowledge published every year [10]. Clinical knowledge may need to be extracted from diverse locations and sources that use different formats. In this regard, many biomedical researchers are looking at developing methods to manage and analyze clinical knowledge in this changeable environment [1,11-13]. One of

the recent technologies that they applied in knowledge acquisition is Semantic Web (SW) technologies [14] to solve the problem of knowledge management, representation, and interoperability of knowledge sources. They have created some Semantic Web-based systems such as COCOON [15], ARTEMIS [16], Semantic-DB [17], Knowledge-Centric Clinical Decision Support Systems [18-19], detecting Alzheimer disease (AD) [20], Semantic-CT [21], sharable CDSS [22], and others. Most existing methods suffer from a lack of a proper mechanism for identifying high-quality knowledge. There exist two main questions: “whether the CDSS contains enough knowledge for diagnosing an unusual disease” and “how to make sure that the knowledge used by CDSS are reliable.”

Regarding the above questions, in this paper, we aim at proposing a semi-automatic approach called Knowledge Quality Assessment (KQA) to discover and assess the clinical knowledge for CDSS.

Research Motivation

The motivation of this research has been inspired by the result retrieved from PubMed search engine. Consider the following query shown in Table 1 which is about “Tuberculosis Arthritis” disease.

Table 1 – The characteristic of a query used in PubMed

[Title/Abstract]	Tuberculosis Arthritis
[Language]	English

The PubMed search engine extracts 18 relevant results for the above query. All of these results are valid, but the question here is “how can practitioners identify the most relevant and accurate result for decision-making process?”. By assessing the abstract/title of the articles, we find out that some knowledge items which are ranked on top of the search results contain little or no useful knowledge.

- 9. [Tuberculosis arthritis. A review of 27 cases.](#)
- 10. Al-Saleh S, Al-Arfaj A, Naddaf H, Memish Z.
Ann Saudi Med. 1998 Jul-Aug;18(4):368-9. No abstract available.
PMID: 17344697
[Similar articles](#)
- 11. [Septic arthritis in patients with human immunodeficiency virus.](#)
- 12. Zalavras CG, Dellamaggiore R, Patzakis MJ, Bava E, Holtom PD.
Clin Orthop Relat Res. 2006 Oct;451:46-9.
PMID: 16906073
[Similar articles](#)
- 13. [Advanced imaging of tuberculosis arthritis.](#)
- 14. Leigh Moore S, Rafii M.
Semin Musculoskelet Radiol. 2003 Jun;7(2):143-53. Review.
PMID: 12920652
[Similar articles](#)

Figure 1- The sample result in PubMed

Figure 1 shows a fragment of the result achieved by the query in PubMed. The “Tuberculosis arthritis: A review of 27 cases” article, which is ranked before the “Advanced imaging of Tuberculosis arthritis”, does not contain an abstract and explanation related to the query. In other words, this article is less useful and has lower quality knowledge compared with “Advanced imaging of Tuberculosis arthritis.” Although one can set up a filter for results only with abstracts, even so, the relevancy does not correspond exactly with the knowledge value of the article. To solve this problem, this research proposes a framework to assess the quality of extracted knowledge for decision making.

Methods

In this section, we first explain the framework of KQA. Then, we describe the candidate metrics for assessing the quality of knowledge. At the end of this section, we also describe a survey, which has been used to rate and validate the importance of candidate metrics.

The KQA framework

Figure 2 shows a detailed view of the KQA framework. In this framework, a query submitted by a user, which represents a knowledge request, will be given to electronic knowledge sources and central knowledge base. Electronic knowledge sources that are used in this project are PubMed, MeSH, and UMLS. They include different clinical knowledge provided through journals, books, and electronic databases, etcetera. The central knowledge base is a machine-readable centralized repository that contains knowledge-based rules extracted from guidelines along with knowledge structure of a particular subject in health domain (e.g. Arthritis). After receiving a query, the KQA system will check the existing knowledge in central knowledge base to find a related result. If the knowledge exists in the knowledge base, the system will deliver the knowledge immediately, if it does not exist, the new knowledge will be extracted from electronic knowledge resources based on query characteristics. The extracted knowledge will be converted to the ontological format and annotated by other information to enrich the knowledge by using Ontology Web Language (OWL). After checking the quality of knowledge by using different quality metrics, the high-quality knowledge will be sent to attach a knowledge quality indicator (KQI) to knowledge item. The KQI indicates the quality of extracted knowledge. Finally, the high-quality knowledge will be used to update the knowledge base.

To update the central knowledge base, the candidate knowledge needs to be checked and approved by the domain experts. This is because it may contradict existing knowledge, and we do not believe at the moment that fully automatic knowledge updating is desirable. However, just being made aware of new highly rated knowledge is an advantage over existing approaches. The approved knowledge will be added to the central knowledge base.

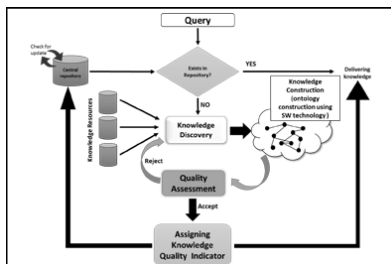


Figure 2 – The KQA framework

Quality assessment metrics

Knowledge quality assessment is a process for checking the quality of extracted knowledge from knowledge sources. Kyoon Yoo et al. [13] noted that knowledge quality should be intrinsically right, contextually relevant, and practically actionable. Based on the Kyoon Yoo model, knowledge quality metrics can be classified into three general categories, including intrinsic, contextual, and actionable metrics. OntoQA [23] is another study which only considered intrinsic metrics. In this paper, we modified the Kyoon Yoo and OntoQA models for categorizing knowledge quality metrics. Table 2 shows three categories of quality metrics proposed in this paper. Based on our categorization, intrinsic metrics are known as the backbone of knowledge. Contextual metrics show how much this knowledge is relevant to a user query. Given a set of actionable metrics indicating that the knowledge is mature, and it can be expanded and adapted for further usage.

Table 2 – Proposed Quality Metrics for KQA

Metric	Description
<i>Age of re-source</i>	The age of knowledge should be declared.
<i>Provenance</i>	The knowledge should be based on valid authority.
<i>Locality</i>	The location of a resource which is relevant to should be declared.
<i>Structure</i>	The knowledge should be in a form that computerized DSS can use, in a consistent structure. E.g. XML. The knowledge should also be rich in concepts (classes) and connections among them (relationships).
<i>Citation</i>	The knowledge should be backed up by accessible citations to research.
<i>Accuracy</i>	How accurate is the knowledge source?
<i>Reliability</i>	The knowledge source will produce the same answer for the same question in different resources.
<i>Relevancy</i>	The resource contains relevant knowledge to support the user query.
<i>Completeness</i>	The answers to queries are complete.
<i>Adoption</i>	The knowledge source gives answers that are useful.
<i>Scalability</i>	The knowledge can be accessed from many systems without losing its meaning.
<i>Timeliness</i>	The resource produces an answer in an appropriate time

Survey for metrics rating and validation

To rate and validate the proposed quality metrics, we conducted a survey among health informatics scholars and practitioners in Health Informatics New Zealand (HiNZ) and the Australasian College of Health Informatics (ACHI) which is also available in [24]. The survey is a questionnaire that has been used for rating the quality metrics through participants. In addition, the participants can propose their own metrics.

Results

We collected the results from 10 experts. Table 3 shows the candidate metrics, which are ranked by participants for knowledge quality assessment for CDSSs.

In this table, the rating is on a scale between 1-5 (1: Not at all Important, 2: Slightly Important, 3: Moderately Important, 4: Quite Important, 5: Extremely Important).

By the survey result, every CDSSs requires an intelligent procedure to check the accuracy, reliability, and relevancy of the extracted knowledge. The accuracy of retrieved knowledge indicates how accurate the knowledge is. It checks the correctness of extracted knowledge against knowledge in the central knowledge base. The reliability metric shows how much the extracted knowledge from different knowledge sources might be similar to each other by using the same query. The relevancy shows how relevant the extracted knowledge is to support the user query. Based on the survey results, in this paper, we focus on developing KQA by assessing accuracy, reliability, and relevancy mentioned in the contextual metrics category shown in Table 3.

As seen in Table 3, the provenance metric has the higher rating average to compare with relevancy. Provenance relates to the perceived reliability of the source of the knowledge. In this research, we put trust on the most well-known knowledge sources for extracting knowledge such as PubMed. There are some metrics that are annotated in the body of extracted knowledge, such as the age of resource, locality, and citation. Such metrics are easy to retrieve and use. The aim of KQA is to check the quality of knowledge before incorporating it in the decision making process. However, there are some metrics (e.g., Adoption, Scalability, and Timeliness) that belong to the actionable category that are related to the quality of knowledge after being incorporated into the decision making process. These may have to be assessed using a study of how knowledge is used operationally in a CDSS.

In the following section, there are some comments collected from participants that identify some metrics that could be useful for further development of KQA.

Person A: *Level of evidence and level of recommendation. This gives flexibility to the CDSS so that it gives more freedom the clinicians. These metrics are found in practice guidelines.*

Person B: *The knowledge is in a form that computerized DSS can use. It is equally important that the knowledge is in a form that the user can use - presentation of information to the user within a CDSS is vital for its safe and effective use.*

Person C: *Validity (the knowledge can be confirmed by using different sources)*

Person D: *Normalization (in the database sense: 3NF). All the ills of denormalized databases are being presented to us as clinicians because database professionals have ignored the importance of normalization.*

Person E: *Weighting. No diagnosis is cast in stone; no observation is 100% "right." At autopsy, 8% to 30% of diagnoses are incorrect. Diagnoses should always be considered to be reputable diagnostic hypotheses. It is important to know how sure a clinician is about an assertion, an affordance not provided by most current EHRs and the like.*

Person F: *Ability to give feedback (to point out possible error or exception)*

Person G: *To me, the structure is NOT just plonking things in XML. It is about the optimal presentation of the minimum of necessary data required for the clinician to do their job. It is difficult, and not well done (as shown in the Epic co-*

trimoxazole incident, and many others besides. Epic may well be better than most).

Person H: *Citations are tricky. It is important that evidence can be traced to its source but not always practicable to include citations in rapid easy to read guidelines.*

Discussion

Based on the results obtained from the survey, we aimed to measure the accuracy, reliability, and relevancy of knowledge that will be used in the CDSS as these are the most highly rated aspects. In this way, we are going to use some SW technology techniques such as ontology matching, ontology similarity, and ontology comparison. However, the process of evaluating results not only relies on proper measures but also user intervention. More precisely, the quality of retrieved results currently will be checked via domain experts. We are currently building a browser to support such knowledge ratings via the SW. A more automated approach that allows crowdsourcing or other approaches may be used in the future. One example may include comparing the outcomes of decisions made on current knowledge with expected outcomes using new knowledge on a database of cases.

In future research, we are going to use PubMed knowledge source to extract knowledge based on Extensible Markup Language (XML) format. We will manually convert the textual information into the ontological-based structure using Protégé ontology editor. We believe that ontological-based structure will be useful for storing knowledge since this structure embrace semantics along with proper annotations. As seen in this research, we assume that textual knowledge is converted to the ontological-based structure. In the future, we would like to develop an approach that automatically converts text structure to ontological structure for further use.

Conclusion

One of the most important activities in healthcare domain is decision making. CDSS can support decision making and may improve patient safety. However extracting up-to-date and high-quality knowledge from the growing mass of knowledge available is difficult and leads to the KA bottleneck. There are many methods and mechanisms for a CDSS to extract and use knowledge to help an expert to make a decision. The CDSS can improve the level of decision making by proposing appropriate knowledge. However, it cannot support how much of the knowledge is accurate, reliable and relevance in the case of comorbidities. Inappropriate knowledge can have negative effects on the decision-making process. Hence, there needs to be a system to check the quality of knowledge used in CDSS to help practitioners make good decisions. This paper aimed to propose a framework for assessing knowledge quality. To validate and rate the candidate quality metrics, we performed a survey among HiNZ and ACHI experts. This has led to a ranking of metrics that we will investigate.

Table 3 – Survey results

Metric	Participants (Rating 1-5)										Average	Total	Count				
	A	B	C	D	E	F	G	H	I	1			2	3	4	5	
	J															N/A	
Accuracy	5	4	5	5	5	4	5	5	5	4	4.7	47				3	7
Reliability	5	4	5	5	5	4	5	5	5	4	4.7	47				3	7
Timeliness	5	3	4	4	3	3	4	3	5	4	3.8	38				4	2
Age of resource	4	2	4	4	5	3	5	3	5	3	3.8	38	1	3	3	3	
Provenance	4	4	4	5	5	4	4	2	5	4	4.1	41			1	6	3
Locality	5	2	2	3	3	2	3	3	4	4	3.1	31				3	4
Relevancy	4	3	5	4	5	3	3	3	5	4	3.9	39				4	3
Completeness	4	2	4	4	5	2	4	4	4	4	3.7	37	2			7	1
Adoption	5	4	5	4	5	2	3	3	5	3	3.9	39			1	3	2
Scalability	5	2	4	4	5	2	4	4	5	3	3.8	38			2	1	4
Citation	4	3	4	5	4	2	3	3	4	4	3.6	36			1	3	5
Structure	N/A	3	4	4	5	3	3	4	4	4	3.4	34				3	5

References

- [1] R.A. Miller, Medical Diagnostic Decision Support Systems—Past, Present, And Future A Threaded Bibliography and Brief Commentary, *Journal of the American Medical Informatics Association* **1** (1994), 8-27.
- [2] D.J. Power, *Decision support systems: concepts and resources for managers*, Greenwood Publishing Group, 2002.
- [3] B. Kotze and B. Brdaroska, Clinical decision support systems in psychiatry in the Information Age, *Australasian Psychiatry* **12** (2004), 361-364.
- [4] W. Bonney, Is it appropriate, or ethical, to use health data collected for the purpose of direct patient care to develop computerized predictive decision support tools, *Stud Health Technol Inform* **143** (2009), 115-121.
- [5] T. O’Kane, J. O’Donoghue, J. Gallagher, A. Aftab, A. Casey, P. Angove, and G. Courtney, MEWS to e-MEWS: From a paper-based to an electronic clinical decision support system, in: *Proceedings of the European Conference on Information Management & Evaluation*, 2010, pp. 301-311.
- [6] B. Leuf, *The Semantic Web: Crafting Infrastructure for Agency*, John Wiley & Sons, Chichester, UK, 2006.
- [7] F. Hayes-Roth, D. Waterman, and D. Lenat, Building expert systems, (1984).
- [8] G. Szulanski, Exploring internal stickiness: Impediments to the transfer of best practice within the firm, *Strategic management journal* **17** (1996), 27-43.
- [9] K. von Michalik, M. Kwiatkowska, and K. Kielan, Application of knowledge-engineering methods in medical knowledge management, in: *Fuzziness and Medicine: Philosophical Reflections and Application Systems in Health Care*, Springer, 2013, pp. 205-214.
- [10] Z. Lu, PubMed and beyond: a survey of web tools for searching biomedical literature, *Database* **2011** (2011), baq036.
- [11] K.-H. Cheung, E. Prud’hommeaux, Y. Wang, and S. Stephens, Semantic Web for Health Care and Life Sciences: a review of the state of the art, *Briefings in bioinformatics* **10** (2009), 111-113.
- [12] K. Sartipi, N.P. Archer, and M.H. Yarmand, Challenges in developing effective clinical decision support systems, *Efficient Decision Support Systems: Practice and Challenges From Current to Future. InTech Open Access Publishing. ISBN 978-953-307-258 6* (2011).
- [13] D. Kyoon Yoo, M.A. Vonderembse, and T. Ragu-Nathan, Knowledge quality: antecedents and consequence in project teams, *Journal of Knowledge Management* **15** (2011), 329-343.
- [14] C. Wroe, Is semantic web technology ready for healthcare, in: *3rd European semantic web conference (ESWC’06), Budva, Montenegro*, 2006.
- [15] E. Della Valle and D. Cerizza, Cocoon glue: a prototype of wsmo discovery engine for the healthcare field, in: *Proceedings of 2nd WSMO Implementation Workshop WIW*, 2005.
- [16] V. Bicer, G.B. Laleci, A. Dogac, and Y. Kabak, Artemis message exchange framework: semantic interoperability of exchanged messages in the healthcare domain, *ACM Sigmod Record* **34** (2005), 71-76.
- [17] C. D. Pierce, D. Booth, C. Ogbuji, C. Deaton, E. Blackstone, and D. Lenat, SemanticDB: A semantic Web infrastructure for clinical research and quality reporting, *Current Bioinformatics* **7** (2012), 267-277.
- [18] S. Hussain, S.R. Abidi, and S.S.R. Abidi, Semantic web framework for knowledge-centric clinical decision support systems, in: *Artificial intelligence in medicine*, Springer, 2007, pp. 451-455.
- [19] H. Mohammadhassanzadeh, W. Van Woensel, S.R. Abidi, and S.S.R. Abidi, A semantic web-based approach to plausible reasoning for improving clinical knowledge engineering, in: *Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on*, IEEE, 2016, pp. 525-528.
- [20] E. Sanchez, C. Toro, E. Carrasco, P. Bonachela, C. Parra, G. Bueno, and F. Guijarro, A Knowledge-based Clinical Decision Support System for the diagnosis of Alzheimer Disease, in: *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, IEEE, 2011, pp. 351-357.
- [21] Z. Huang, A. ten Teije, and F. van Harmelen, SemanticCT: A Semantically-Enabled System for Clinical Trials, in: *Process Support and Knowledge Representation in Health Care*, Springer, 2013, pp. 11-25.
- [22] Y.-F. Zhang, L. Gou, Y. Tian, T.-C. Li, M. Zhang, and J.-S. Li, Design and Development of a Sharable Clinical Decision Support System Based on a Semantic Web Service Framework, *Journal of medical systems* **40** (2016), 1-14.
- [23] S. Tartir, I.B. Arpinar, M. Moore, A.P. Sheth, and B. Aleman-Meza, OntoQA: Metric-based ontology quality analysis, (2005).
- [24] S. Zolhavarieh and D. Parry, A study to validate metrics for quality of clinical knowledge in Clinical Decision Support System, in: *New Zealand*, 2015.

Address for correspondence

Seyedjamal Zolhavarieh
 Department of Computer Science
 Auckland University of Technology
 Private Bag 92006
 Auckland 1142
 New Zealand
 Email: szolhava@aut.ac.nz

V. Quality and Safety, and Patient Outcomes

This page intentionally left blank

An Innovative Approach to Jointly Scheduling and Assigning a Consultation Time to Patients Arriving in the Emergency Department

Faten Ajmi^a, Sarah Ben Othman^a, Hayfa Zgaya^b, Slim Hammadi^a, Jean-Marie Renard^b

^a Cristal Laboratory UMR 9189, Ecole Centrale de Lille, Villeneuve d'Ascq, France

^b EA2694 University Lille 2, Lille, France

Abstract

Emergency departments (ED) are facing problems related to the growing demand of care. Patients' management is carried out according to the type of patient and care required: already scheduled patients and non-scheduled urgent and non-urgent patients arriving in the ED. One of the main problems confronted in hospitals is the permanent interference between these different types of patients to be treated under the stochastic behaviors of consultation time and arrival flows, which prevents any prior planning. The present work proposes a dynamic scheduling method, considering the impact of new patients' arrivals on the treatment of patients already scheduled to minimize the mean waiting time of patients in the ED. The originality of this work is to assign, at the time of arrival, a scheduled time to each patient in order to reduce their stress. The performance of the proposed method is examined through a concrete application in the Pediatric Emergency Department of CHRU of Lille.

Keywords:

Health care rationing, Emergencies, Waiting lists.

Introduction

Nowadays, hospital Emergency Departments (EDs) have a strategic place in modern health care systems and represent the main gateway to the hospital. This key role is expected to strengthen in future years due to the steady growth in the number of arrivals and the increasingly demands of patients [1]. These changes set many problems for different actors in public health, including problems of functioning to cope with this increase in consultations, and a non-negligible healthcare spending [2-6]. However, while the majority of healthcare systems in the world are confronted with this reality, the ways to face them differ from one country to another [7, 8]. In order to better understand current problems and challenges, it is first necessary to place emergency services in their general context in order to understand their functioning and their specific characteristics.

Indeed, the question in France is still topical, and hospital emergencies are taken into account at the highest level within the national organizations. The Code of Public Health specifies all the missions of these services. It is a mission characterized by:

- The reception: to accept every patient arriving in an emergency, including a psychiatric emergency, 24 hours a day, any day of the year, and take care of him, especially in case of distress and vital emergency.
- Translation, definition of demand, needs qualification of the patient with a symptom analysis work.

- The preservation of life, stabilization of a condition thanks to the care, and "urgent diagnostic and/or therapeutic procedures", adapted to the situation.
- Patient orientation at the right time, in the right sector, by the appropriate services.
- Short-term hospitalization in certain cases, depending on the resources and the reception capacity, and according to the functioning of the health care organization.

The care of an emergency patient mobilizes different resources of health organizations: upstream of the hospital, within the health establishment and downstream of the health facility (other institutions). However, the notion of emergency in the medical field remains a source of great ambiguity. In a literature review, it is not difficult to see that if there is a consensus, it is summed up in the fact that an "emergency" requires an action, and an immediate decision. In case of emergency, it is essential to intervene immediately irrespective of the type of emergency and that it is necessary to respond to any request for care, even if the latter is non-programmed and is not a source of life-threatening distress, regardless of the seriousness of the problem.

A situation is called "urgent" because it is considered serious and it calls for a quick response. It is often subjective and intimately linked to the value system of the person who states the emergency judgment. This means that between the perception of urgency by the patient (perceived urgency) and the medical definition of emergency by a medical staff (actual emergency), the gap is often important. A gap may be even greater than the urgency of a case considered in relation to the urgency of other cases [9].

Ensuring patients' satisfaction in the ED is part of a continuous improvement strategy [10]. The length of the consultation, waiting time, dissemination of information, and interaction with practitioners in the ED have the strongest predictors of patient satisfaction. This does not mean that patients are not aware of the necessary treatment time. But they need to be well informed about the progress of treatments on one hand, and on the overall progression of care on the other hand. In addition, it is so important to give a provisional time of consultation to each patient to reassure him. This also reflects the level of overcrowding, as the length of stay of patients appears to increase considerably when the number of patients increases in the ED. Nevertheless, it is difficult to assess patient satisfaction because of the complex interaction between expectations, perceptions, and reality.

In this paper, we take into account patient waiting time at different stages of patient journey: waiting time between the registration and the provisional time of consultation called "scheduled time", and between scheduled time and the time of

the first consultation. Waiting times can be the result of the bottlenecks in the ED, or the arrival of more urgent cases.

Our objective is to propose an innovative scheduling approach able to solve the problem of the interference between already scheduled patients and non-scheduled urgent and non-urgent arriving patients to emergencies. We aim to optimize patients' treatment process, and minimize waiting time in the ED. We take into account the availability of medical staff members belonging to different medical teams.

Methods

Scheduling environment

Assumptions

- There is a medical staff member engaged in the scheduling horizon in an ED. In a scheduling horizon, the number of already scheduled patients is N_s , while the expected number of non-scheduled patients is N_{ns} . All the non-scheduled patients arrive randomly at the ED, then a scheduling time must be assigned to them at the time of their arrival.
- In France, the ED is always opened and each arrived patient p should be registered in the reception desk at time trp . All patients who arrive at the ED are accepted, and should be treated in the current or the next horizon scheduling.
- Each patient corresponds to a set of health care operations to be executed in a parallel or in a sequential manner by one or more medical staff members (physician, nurse, intern, etc.).
- Medical staff members are organized in teams. Each team contains at least one physician. We can have more medical staff members (nurse, pediatrician, etc.) in one medical team depending on the patient pathology.
- The scheduling horizon H starts at time D_H and ends at the time F_H . In this paper, we consider that the duration of one horizon is 4 hours.
- The scheduling horizon is divided into several periods with different length, not necessarily the same. If two periods have the same length, the number and the duration of slots in each period may differ, and generally one period contains multiple slots. A slot is allocated to one scheduled patient. Each period contains at least one slot. The scheduled time of a slot is given by the starting time of the period to which it belongs. Therefore, if two or more slots are included in a period, the scheduled patients assigned to the same slot have the same scheduled time.
- Under the stochastic behavior of consultation time of medical staff, let $C_{m,o}$ be the average consultation time of the medical staff member m having the skill o to treat patient j .
- When the medical staff member becomes available, the waiting patient with the earliest scheduled time is called. If the waiting room is full, and there is no possibility to call all the patients in the same scheduling horizon, the remaining patients and the new arriving patients will receive a schedule time in the next scheduling horizon.
- In ED, priority is given to the most urgent cases. So, at the arrival of urgent patients, the current scheduling can be interrupted and a rescheduling is necessary,

because these patients should not wait for consultation.

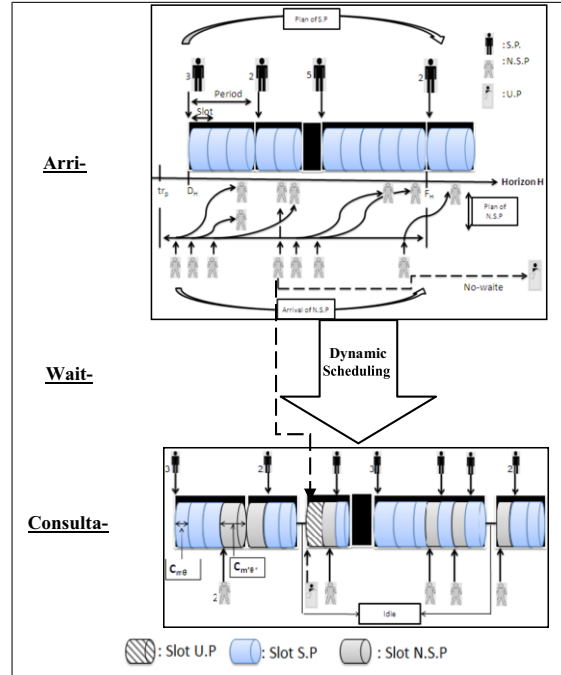


Figure 1 – Scheduling Environment

Performance measures

Let the waiting time of scheduled patient P_j , $W_{s,j}$ be the sum of patients' waiting time between the registration and the given scheduled time W_{ar} and the waiting time before the first consultation W_{fc} , where:

$$W_{s,j} = W_{ar,j} + W_{fc,j} \quad (1)$$

$$W_{ar,j} = \max(0, t_{ar,j} - t_{s,j})$$

$$W_{fc,j} = \max(0, t_{fc,j} - t_{s,j})$$

Where $t_{ar,j}$, $t_{s,j}$ and $t_{fc,j}$ represent, respectively, the arrival time, the scheduling time, and the first consultation time of the patient j .

The waiting time of non-scheduled patient P_k , $W_{ns,k}$ be the sum of patients' waiting time between the registration and the given scheduled time $W_{ar,k}$, the waiting time before the first consultation $W_{fc,k}$, where:

$$W_{ns,k} = W_{ar,k} + W_{fc,k} \quad (2)$$

$$W_{ar,k} = \max(0, t_{ar,k} - t_{s,k})$$

$$W_{fc,k} = \max(0, t_{fc,k} - t_{s,k})$$

The two equations (1) and (2) are mathematically equivalent, but semantically different. In fact, the present study proposes a scheduling method that assigns a scheduled time to each non-scheduled patient at his time of arrival, and guides him to go to the waiting room at his scheduled time. Thanks to the assigned scheduled time, W_{ar} is calculated for each registered patient. The objective is to comfort patients and reduce their stress by giving them, in advance, a waiting time till the first consultation. If the first consultation time is equal to the scheduled time, then the waiting time $W_{fc,j}$ and $W_{fc,k}$ are equal to 0. In case of perturbation (overcrowding situation, absence of medical staff, worsening of patient health state, etc.), a rescheduling is done, and the first time of consultation increases, which makes the waiting time longer for the patient.

Usually most of patients prefer earlier scheduled times, especially at their arrival time. To respect these preferences, the waiting time based on the arrival time War_k should be reduced, by assigning as quickly as possible an available medical staff with an appropriate skill.

In this paper our objective is to minimize the total waiting time W of both scheduled and non-scheduled patients, by taking into account, dynamically, the availability of medical staff members and their skills.

$$W = \min \sum_{j=1}^{Ns} \sum_{k=1}^{Nns} (Ws, j + Wns, k)$$

We choose to adopt an aggregative approach without handling appropriate weights to have optimizing solutions. In fact, in real life health care situations, it is very difficult to define suitable weights for these criteria. This work, studies the results of simulations generating some of these criteria separately or some of them jointly.

Scheduling method

Characteristics of non-scheduling patients

As it is mentioned above, the driven idea is to have a scheduling method realizing the shorter waiting time for both scheduled and non-scheduled patient. In order to reduce the effect of the stochastic aspect of arrival flow, we assign to each non-scheduling patient a scheduled time at his arrival time.

The present study gives good solutions by focusing on the similarity between the job-shop scheduling in the flexible manufacturing systems, where the workload is an important factor for the decision maker. In the complex manufacturing system, minimizing the total workload allows to avoid the bottleneck in production system generating a reduction waiting time.

Under the condition of the present work, the Health-Workload $HW_{p,H}$ of each period p in the scheduled horizon H , can be expressed by the average consultation time multiplied by the number of scheduled patients in the period. This means that once scheduled patents are set in the period. The scheduling time for non-scheduled patients is decided at the moment of their arrival in the ED by calculating the expected $HW_{p,H}$ that includes them in the candidate periods. If a maximum value of $HW_{p,H}$ is reached, then the next period becomes the candidate

By selecting an adequate solution for scheduling patient that allocates a sufficient initial $HW_{p,H}$ to each period p and by assigning each non-scheduled patient to an appropriate period in order to get a reasonable $HW_{p,H}$, the conflicting criterion W is balanced. Controlling and keeping the total $HW_{p,H}$ for each period p and each horizon H below the specified maximum value $HW_{p,H}$, will complete shorter waiting times of patients as well as shorter total idle times of medical staff members.

Dynamic scheduling method of non-scheduled patients

This work assumes that the scheduling horizon, which represents consultation time window, is divided into several periods as already mentioned above. To calculate a scheduled time of a non-scheduling patient, we use a first empty slot in the period as shown in Figure 2.

In principle, the start time of the first empty slot gives the scheduled time of the patient. The maximum number of non-scheduled patients accepted in the period is difficult to estimate, because slots have different lengths in the same period of them jointly.

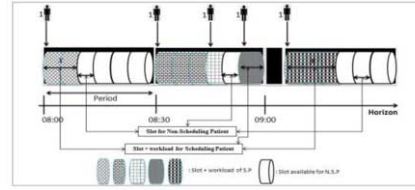


Figure 2 – Example of determining patient scheduling time based on free slots

Algorithm: Search-First-Free_Slot (Period p , Horizon H , Patient w , Deb_p , Fin_p)

Inputs:
 Nbs_p : Number of slots in the period p .
 TAB [1... Nbs_p] a table contains the length of each slot.
 $FREE$ [1... Nbs_p] a table contains 1 or 0.
 $FREE(i) = 1$ the slot i is free else the slot i is full.

Output:
 T_w the start time of the first free slot

Begin

Calculate $HW_{p,H}$;

If ($HW_{p,H} + C_{m,0}^w < HW_{p,H,max}$) **then**

begin

Assign-patient- w to the period p ;

$T = Deb_p$: the slot starting time

For $i = 1$ to Nbs_p **do**

If $FREE(i) == 1$ **then return** T

Else $T = T + TAB[i]$

End_For

IF ($T == Deb_p$ and $i == Nbs_p$) **then**

begin

$T = Fin_p$

Return T

End_If

Else return the period p is overloaded

End_Search-First-Free_Slot

Algorithm: Scheduling_arrived_patient (Patient w , Time t , Horizon H)

Input:

Deb_H : Starting time of the horizon H represents the starting time of consultation;

t : Current arrival time of patients

Fin_H : The ending time of the horizon H ;

$NB_{p,H}$: Number of periods in the horizon H ;

TAB_H : Table [1... $NB_{p,H}$] contains the different lengths of each period.

Output:

T_w the starting time of the first consultation specifying Horizon, period and slot.

Begin

If Urgent_patient **then** No-wait-consultation

If $t \leq Deb_H$ **then** $p = 1$ (the first period in the horizon)

Else

For ($i = 1$ to $NB_{p,H}$) **do**

If $t \leq Deb_H + TAB_H[i]$ **then**

$p = i$;

Save the starting time of the period Deb_p

Save the ending time of the period Fin_p

End_If

End_For

End_If

$T_w = \text{Search-First-Free_Slot}(p, \text{Horizon } H, \text{Patient } w, Deb_p,$

$Fin_p)$

End_Search-First-Free_Slot

The ending time of the consultation window $t_{max,H}$ is used for the scheduled time for patients who are not included in any of the periods in the horizon H . The present study assumes that each period p may have its own length Δp , and the starting time of each period is x minutes behind the starting time of its first free slot. The maximum workload level per period is $HW_{p,H,max}$. For example, if there are 3 scheduled patients j and 2 non-scheduled patients k in period p , then $HW_{p,H}$ is given by $HW_{p,H} = 3 * C_{m,0} + 2 * C_{m',0}$.

Simulation and Results

For simulations, we did observations in the Pediatric Emergency Department of CHRU of Lille, our partner in the

project ANR HOST. We have a 3-years database, 2011-2013, characterized by overcrowding and pics of activity.

Figure 3– Database of the PED of CHRU of Lille

Currently, there is no decision support system or information system able to manage the PED and the problems corresponding to an overcrowding situation. Medical staff members give the highest consultation priority to the most urgent patients and $HW_{p,H,max}$ the already scheduled patients. Non-scheduled patients in the PED have to wait in the waiting room, and some- times in corridors, without obtaining a scheduling time for the first consultation, which increases their stress.

Our proposed approach performance is tested by comparing it to the current traditional method used in the PED. The database analysis allowed us to define the evolution of patients' waiting time which sometimes seems to be excessive.

Experimental conditions

From the database, we choose an overcrowded day (25th of February 2013) divided into 6 horizons, the duration of each one is 4 hours. Each horizon represents the consultation time window.

Table 1– The average values under different maximum workload levels

$HW_{p,H,max}$ (min)	Number of urgent patients with different pathologies	Scheduled patients		Non-scheduled patients		Total idle time (min)	Mean time of no-wait consultations (min)
		\bar{W}_{ar} (min)	\bar{W}_{fc}^s (min)	\bar{W}_{ar} (min)	\bar{W}_{fc}^n (min)		
120	2	21,71	10,20	27,89	12,21	11,20	20,28
	4	21,29	9,80	24,17	12,94	10,40	25,40
	6	21,00	9,56	20,32	13,20	9,20	26,20
	8	20,90	8,94	27,10	15,37	8,30	32,01
110	2	19,80	9,76	26,58	14,62	9,10	25,50
	4	19,10	9,20	26,23	16,50	8,88	24,75
	6	18,65	8,70	28,10	17,02	8,00	22,00
	8	18,25	8,52	30,00	18,43	7,45	27,50
100	2	20,24	12,40	28,54	15,60	4,80	30,00
	4	21,00	13,43	29,00	16,83	5,24	32,42
	6	23,89	15,17	29,50	18,15	5,50	35,79
	8	26,27	15,50	31,12	18,70	6,10	40,31

For simulations, we consider that the periods and the slots have different lengths. Each consultation time window is composed of 5 periods. The length of each one is alternatively 1 hour and 30 min. The one-hour period has 8 slots whose length is alternatively 5 min and 10 min. The 30 min period is composed by slots of 5 min and 10 min. We consider for the scenario that the medical staff is represented by only physicians in this numerical example. The mean consultation time of scheduled patients $C_{m,o}^s$ is 5 min, and the mean consultation of non-scheduled patients is $C_{m,o}^n$ 10 min. The arrival of non-scheduled patients follows a Poisson distribution. The mean inter-arrival time of non-scheduled patients is 10 min. So, we have 24 arriving patients to the PED in one horizon. Table 1 summarizes the average values under different values of $HW_{p,H,max}$ (min) .

According to the table, when $HW_{p,H,max}=100$ min, the average waiting times of scheduled and non-scheduled patients have increased. The idle time of physicians has decreased. This is due to the arrival of urgent patients to the PED. In fact, these patients have complex pathologies which need more consultation time.

This phenomenon proves the impact of the interferences between scheduled, non-scheduled, and urgent non-scheduled patients. These interferences depend on the number of incoming patients and their pathologies.

Figure 3 represents the relationship between the number of no-wait consultations and the total idle of physicians.

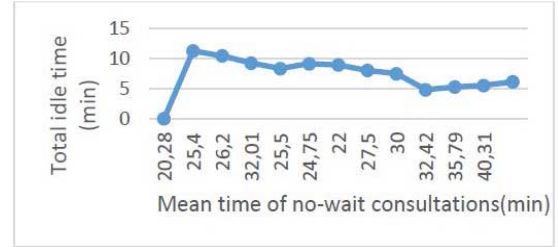


Figure 4 – The relationship between the number of no-wait consultations and the total idle of physicians

Through figure 4, we notice that there is sensitivity between the number of urgent patients and the different waiting times. The mean time of no-wait consultations is inversely proportional to the total idle time of the physician. This also indicates that, while adjusting the total idle time of physicians, the frequency of no-wait consultations is minimized. In fact, when the pathology of urgent patients is complicated to treat, the idle time of physicians becomes smaller. In this case, scheduled patients in the next periods are able to receive earlier consultation which therefore decreases the value of W_{ar} .

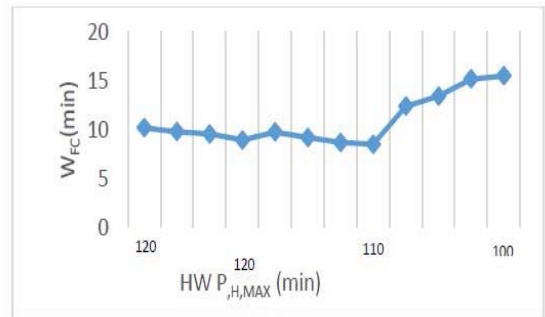


Figure 5 – The relationship between the waiting time before the first consultation and the maximum health workload

When the maximum workload $HW_{p,H,max}$ in the period p for the horizon H is reduced, it leads to reduce the scheduled times of scheduled patients and non-scheduled patients. The increase of $HW_{p,H,max}$ leads to longer waiting times for both scheduled and non-scheduled patients (see figure 5).

Discussion

In the PED of CHRU of Lille, we noticed that patients are often dissatisfied, many feel that their waiting is excessive and that their care should be immediate. This is also a matter of subjectivity, or of poor ranking in the order of priorities of patients. Because, it should not be forgotten that in these

circumstances priority is given to patients in vital emergencies. In sum, the main cause of overcrowding situations is that the demand exceeds the supply generated by the functioning nature of this structure, because the notion of appointments does not exist. This limits the margin of organization and planning. The situation becomes even worse, when several patients arrive at the same time to the PED.

Usually, in the waiting room, patient with the earliest scheduled time is called first for consultation. However, under the stochastic and irregular consultation time and arrival of more urgent patients, determining a method for assigning a scheduled time to each patient arriving is a research issue. Perturbations that can occur during the treatment of patients already scheduled are able to modify the scheduling. A considerable research effort has been devoted to scheduling problems related to emergencies. Although the arrival of new patients and the appearance of urgent and non-scheduled patients produces a complicated situation in patients' scheduling, it seems that this kind of scheduling problem attracts little attention in the literature.

Our main consideration in this work is the development of dynamic scheduling solutions with rolling horizon to simplify the work of medical staff and minimize the average waiting time in the PED for both scheduled and non-scheduled patients. Through assigning a scheduled time to patients, medical staff members can have an idea about their workload during the horizon and patients are reassured. The proposed approach is also able to adapt the available human resources in the PED in order to afford a good quality of health care for patients. In fact, the arrival of urgent patients in the case of an overcrowding situations leads to the minimization of the total idle time of medical staff. Our approach is based on the dynamic and reactive generation of treatment plans. The information and especially the treatment timestamps are given through a dynamic orchestration workflow architecture modeling the patient journey through agents [11].

Conclusions

In this paper, we have proposed a dynamic scheduled system with rolling horizon, by assigning a scheduling time to each patient at the time of his arrival. Our innovative scheduling approach is able to solve the problem of the interference between already scheduled patients and non-scheduled urgent and non-urgent patients arriving in the ED. Our target is to optimize the treatment process of patients, and minimize their waiting time in the ED. For future work, we aim to improve this approach through an alliance with genetic algorithms and multi-agent systems in order to solve the problem of multi-skill health care tasks scheduling in the ED.

References

- [1] M. Cooke, J. Fisher, J. Dale, E. McLeod, A. Szczepura, P. Walley, and S. Wilson. Reducing Attendances and Waits in Emergency Departments. A systematic review of present innovations, Report to the National Coordinating Centre for NHS Service Delivery and Organization R & D (NCCSDO), 2004. Available at: <http://www.sdo.nihr.ac.uk/files/project/29-finalreport.pdf>
- [2] C. Benjamin, M.D. Sun, M.P.P., Renee, Y. Hsia, M.D. Robert, E. Weiss, and et al., Effect of Emergency Department Crowding on Outcomes of Admitted Patients, *Annals of Emergency Medicine* **61**(6) (2013), 605-611.
- [3] R.W.T. Derle and J.R. Richards, Overcrowding in the nation's emergency departments: complex causes and disturbing effects, *Annals of Emergency Medicine* **35**(1) (2000), 63-68.
- [4] P.M. Pines, J.A. Hilton, E.J. Weber, A.J. Alkemade, H.A. Shabanah, P.D. Anderson, M. Bernhard, A. Bertini, A. Gries, S. Ferrandiz, V.A. Kumar, V.P. Harjola, B. Hogan, B. Madsen, S. Mason, G. Ohlén, T. Rainer, N.

- Rathlev, E. Revue, D. Richardson, M. Sattarian, and M.J. Schull, International perspectives on emergency department crowding, *Academic Emergency Medicine* **18**(12) (2011), 1358-1370.
- [5] J.D. Schuur, A.K. Venkatesh, The growing role of emergency departments in hospital admissions, *New England Journal of Medicine* **367**(5) (2012), 391-393.
- [6] L. Nathan, M.D. Timm, L. Mona, M.S. Ho, W. Joseph, M.D. Luria, Pediatric Emergency Department Overcrowding and Impact on Patient Flow Outcomes, Issue, *Academic Emergency Medicine* **15**(9) (2008), 832-837.
- [7] F. Galvis-Narinos and A. Montélimard, Le système de santé des États Unis, *Pratiques et Organisation des Soins* **40**(4) (2009), 309-315.
- [8] W. Renaud, D. Désir, and F. Hut, La réforme Obama du système américain de soins de santé, *Courrier hebdomadaire du CRISP* **35-36**(n° 2160-2161) (2012), 1-100.
- [9] [SFMU, 2008b] Société Française de Médecine d'Urgence. Les français et l'accès aux soins urgents non programmés. 2ème Congrès de la SFMU, Paris, June 2008.
- [10] W. Chandoul, H. Camus, N. Zoghalmi, S. Hammadi, and A. Martinot, A robust assessment of effective healthcare demand in the Pediatric Emergency Department, In: 2014 International Conference on Control, Decision and Information Technologies (CoDIT), 2014, 195-200.
- [11] S.B. Othman, H. Zgaya, S. Hammadi, A. Quilliot, A. Martinot, and J.M. Renard, "Agents Endowed with Uncertainty Management Behaviors to Solve a Multiskill Healthcare Task Scheduling". *Journal of Biomedical Informatics* **64** (2016), 25-43.

Address for correspondence

Sarah Ben Othman, CRISTAL UMR CNRS 9189, Ecole Centrale of Lille, Avenue Paul Langevin, Cité scientifique, 59650 Villeneuve d'Ascq, France, e-mail: sara.ben-othman@ec-lille.fr

Design of a Mobile Application for Transfusion Medicine

MA Albornoz, S Márquez, L Rubin, D Luna

Departamento de Informática en Salud, Hospital Italiano de Buenos Aires, Argentina

Abstract

One of the most frequent error in transfusion medicine is the failure in verifying the patient's identity prior to transfusion. This paper describes the design and development of a Mobile Application (MA) for transfusion medicine. The app uses barcode and QR reading technology for the verification of the patient's identity and the administration of blood components when making a blood transfusion. Physicians, developers, technicians of transfusion medicine and a User Centered Design team participated in the design. The inclusion of end users was fundamental to get full representativeness of their workflow. The project was based on agile methodologies of project management and software development.

Keywords:

Blood Transfusion, Patient Identification Systems, Patient Safety

Introduction

Today transfusion medicine has become a safe and essential practice in hospitals, but it is a process in which the slightest mistake has the potential of causing great morbidity or even death. A study in New York estimated that 1 in 12,000 transfusions are caused by ABO compatibility mistakes. The Serious Hazards of Transfusion (SHOT) reports in England showed that the risk of death from transfusion in 2012 was 1 in 322,580 transfused components and a had a morbidity greater than 1 in 21,413 [1, 2]. Approximately 70% of incorrect blood components administration occurs in clinical areas. The failure of the final check of the patient's ID bedside is one of the most frequent error. The patient's identity verification is a critical point to provide security [3, 4, 5, 6, 7].

The use of new technologies is one of the best strategies to avoid this kind of mistakes. The barcoding technology allows increasing patient safety with the verification of their identity at the time of taking a new sample and administering a blood component. Like this, it becomes a physical barrier that allows errors reduction in the stages of the circuit and the improvement of the product's identification and correct patients rate [8, 9, 10, 11].

Currently, several papers describe the creation and use of barcode systems with Personal Digital Assistant (PDA)

systems, but no published works were observed allowing the use of any Smartphone or tablet device with Android operating systems (OS) in the area of transfusion medicine [8, 9, 10, 11].

The purpose of this paper is to describe the iterative design and development of a Mobile Application (MA) that will be supported in Android OS, which uses bar code and QR reading technology, for the safe administration of blood components when a blood transfusion is performed. We will describe the current transfusion process in our institution and then the one addressed by the MA.

Methods

Setting

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit healthcare academic center founded in 1853, with over 2,700 physicians, 2,700 other health team members (including 1,200 nurses) and 1,800 administrative and support employees. The HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires city and its suburban area. It has a Health Maintenance Organization (Plan de Salud) that covers more than 150,000 people and also provides health services to another 1,500,000 people who are covered by affiliated insurers. Between 2013 and 2014, over 45,000 inpatients were admitted to its hospitals, there were 45,000 surgical procedures (50% ambulatory) and 3,000,000 outpatient visits. In addition, the HIBA is a teaching hospital, with over 30 medical residency-training programs and 34 fellowship programs. There are currently 400 residents and fellows in training.

Since 1998, the HIBA runs an in-house developed health information system, which includes clinical and administrative data. Its Electronic Health Record (EHR) system, Italice, is an integrated, modular, problem oriented and patient centered system that works in different clinical settings (outpatient, inpatient, emergency and home care). Italice allows computer physician order entry for medications and medical tests, and storage and retrieval of tests results, including images through a picture archiving and communication system. It has been recently certified by the HIMSS as level 6+ in the Electronic Medical Record

Adoption Model, being the first hospital in Argentina and the second in Latin America reaching this stage.

The HIBA Transfusion Medicine Service (TMS) is composed by 7 staff physicians and more than 30 transfusion medicine technicians (TMT), who works distributed in 3 shifts. At each shift, 4 technicians make rounds through the hospital performing the requested procedures, while others support the process in the MTS. They carry out an approximate of 110 transfusions of hemocomponents daily, and receive approximately 50 blood donors daily. For the registration of donors, recipients and transfusions they use an in-house system called HEMOTRANS®. It does not currently interoperate with the EHR, although its developers along with the Department of Health Informatics are working on an integration plan. The HEMOTRANS® system is responsible for crossmatching the information between patient's samples and donors. It also stores hemoproducts intake and generates reports.

From October 2015 to November 2016, 38,663 transfusions were performed in HIBA. This is equivalent to an average of 3221 monthly procedures and 110 daily. In that period, 15 transfusions adverse reactions (present within the initial 15 minutes after the transfusion started) were recorded in structured entry forms within the HCE.

In order to understand the current circuit of transfusion medicine, meetings were held in a multidisciplinary team conformed by health informatic physicians, MTS staff, TMTs and user centered designers. The current transfusion process and the steps to be taken to address the design of future application were agreed. Prior to submitting the request to the developers, test mockups were performed and iteratively modified with the TMTs.

The current transfusion process:

1. The physicians make the transfusion request from the EHR. An single request may contain different types of practices (for example 2 units red blood cells and 7 units of platelets). The order also has the emergency status of the patient, and a summary of the hematologic status of the patient.
2. The TMS coordinators print all the transfusion requests and distribute them among the technicians according to their availability. They print the informed consent and barcode labels with patient identification from the EHR and verify in HEMOTRANS® if they have already been grouped within 72 hours prior to the transfusion to be performed. If this is the case, they use these stored samples to group the donor's blood. When there is not a registered sample, a blood extraction is made. This is used for serological follow-up is and blood grouping to corroborate the group and factor before the transfusion is done. Each tube is identified with the respective patient identification label. The results are then entered into the system and in the EHR.
3. In the TMS, the blood product bag is selected from the blood bank, and barcoding is performed on it from the HEMOTRANS®, in order to subtract the

product from the stock and to check its compatibility with the patient's blood to be transfused.

4. The next step happens in the patient bedside. His identity (name and date of birth or identification bracelet against the paper order) is checked verbally and if they're correct the blood transfusion is performed.
5. If there is a transfusion reaction in the first 15 minutes, the TMT registers it in the EHR and the medical team is notified.
6. The last step happens when the TMT closes the order from a desktop computer generating an automatic progression note with all the transfusion process' data.

Results

Design and development of the MA

In August 2016, a team of resident physicians from the Health Informatics Service with TMT and doctors from the TMS defined the steps and activities that formed the process of a transfusion.

The MA was designed by a team of health informatics physicians, software developers and a user centered design (UCD) team, with the insights of transfusion medicine physicians and technicians.

The project was based on agile methodologies of project management and software development. Subsequently, weekly personal interviews were conducted with 10 technicians to deal with the details of the process. They were accompanied on their daily visit and the process of realizing several transfusions was observed to understand their complexity and to detect problems and opportunities of future implementation.

Layout Instance

The health informatics team generated a Project Charter document indicating the project scope and magnitude, and a Gantt chart setting the project time frame. Through the software Balsamiq® MA several mockups series were generated.

Mobile application prototyping instance

The mockups were tested with the TMT, and modified iteratively according to the tests results. The first series of mockups were in low fidelity, the followings improved their quality. With the mockup final version, a PDF document (portable document format) was designed, emulating the workflow and functionality of the final application, forming a high-quality test application.

Tests of MA

This instance allowed users to test the AM with a smartphone before starting its development. To this end, different use cases scenarios were created with specific tasks to be performed in TMS. The TMTs that were involved in the design process of the MA were not the same than those who

tested the final version, to avoid any bias. 5 TMTs were gathered and a smartphone was given to them, along with the task to be performed. The TMTs successfully completed the process with very good acceptance. The design team filmed them, with their consent, to obtain more qualitative conclusions later.

Development Instance

Once the TMTs and sponsor approval was obtained, a development request was made to the development team who are currently completing the web application developed with Ionic Frameworks v1 and AngularJS v1.3. Material was used as a template (extension of the Ionic Framework) [12, 13].

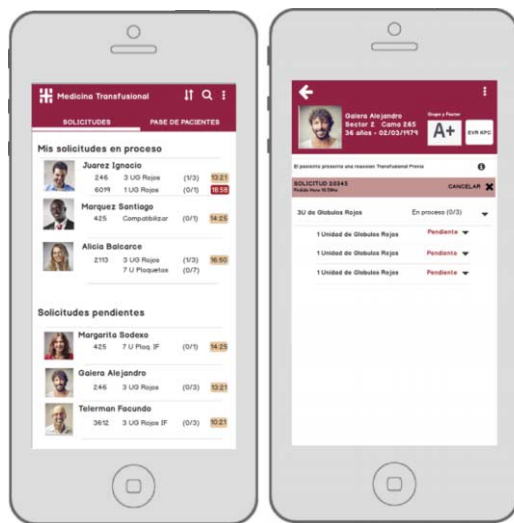


Figure 1 - Main menu and patient space

Workflow in the mobile application

The TMT logs into the AM by entering username and password. When he starts his work shift he can view the worklist of all the products requested in the EHR, without the need to print any order on paper. He selects and assigns to his own worklist the requests in with he will work. If there are any pending procedures from another work shift, there is a "patient handoff" where the TMT takes the patient and associates him to his own work list, with the purpose of allow continuity of care. The list is configurable by request creation's hour, request urgency and sector from where it was generated. The requests can be grouped by patient, if there are several of them for the same patient.

As can be seen in Figure 1, when the TMT clicks on a request on the list, the next screen of the app is the "Patient Space". On its header contains patient's name, date of birth, hospital location, allergies, infectious isolation, blood type, updated personal photo, among others. The "Patient Space" body

contains the requests and the different actions that can be made through the app.

Depending on the request, the MA's main actions are:

- **Sample:** This option allows the technician to perform barcoding on the patient's identification bracelet and on the label of the newly extracted sample tube by checking that the labels are on the correct patient. This blood grouping is serologically monitored and a new corroboration of group and factor, genotype and detection antibodies is made prior to transfusion. (Point 2 of the process described above).
- **Transfusion:** This option enables a blood product transfusion. It needs the identification of the patient, the request order and the product to be transfused. The AM can read any type of barcode identification or QR codes. If there is any data inconsistency at the time of the barcoding the application gives an alert message and registers it. Once the first 50 ml of the transfusion are in the patient bloodstream, the TMT registers if there was any transfusion reaction and generates an automatic progression note from the application directly into the EMR. (Point 4 to 6 of the process described above)
- **Deliver:** The transfusion medicine technician transfers the hemoproduct, for another technician or nurse to perform the transfusion later. These cases are usually given in the operating room or in the neonatal care unitPostpone: It happens in any situation when the transfusion can not be performed. It allows the TMT to register this in the. Before the AM this was recorded with a manuscript evolution in the EHR.

Bedside, according to the circuit that is being performed, the AM indicates the steps to complete. The AM allows the tracking of the entire process and provides the possibility to see the previously performed procedures, report adverse reactions and generate a direct record in the EHR. By having instant communication via Wi-Fi each of the procedures impact in the form of CDA (Clinical Document Architecture)[14] and allows its visualization to the other members of the health team. In this way the process was complete without the need to print paper orders.

Discussion

This paper describes the design and development of an AM that will allow to assist the MTS in their process. It is impossible to create a representative tool without the users participation involved in the task. The user centered design area is increasingly involved in AMs production, trying to make it easier to learn, simpler to use, and gain a good acceptance and perception of utility by users[15][16][17]. This design was made by constant and iterative work with end users of the application. The design team aim was for the system to be able to achieve the goals set with a balance

between systematization and flexibility. We achieved a good level of satisfaction for each of the TMT participants with whom we worked. The field of health is gradually adopting user centered design strategies so that computer applications match the tasks or activities for which they were designed [18].

According to literature, we need to make more and more interactive designs with multiple versions, tasks and environments, including the user at the beginning of the project and then carry out product evaluations aiming to improve quality, effectiveness and efficiency [19].

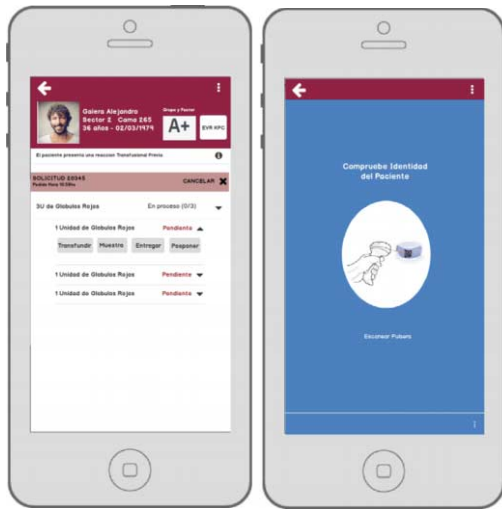


Figure 2 - Main actions in the "Patient Space" and "Barcoding of an identifying bracelet"

The application will be responsive for both smartphones and tablets that use Android OS. The application will be available within a package of functionalities that are being designed in our institution. Currently we have an MAs already implemented for the patient's transfer by the stretchers who perform barcoding on the patient's bracelet and the place of destination where he has to be taken. Another MA is also being completed, which nurses will use to check the "5 correct" when giving a medication, taking vital signs of the patient and other activities performed by the nursing area. This set of applications interact with each other, generating greater communication in the health team, better continuity of care and a quick online update in the EHR.

Limitations

The transfusion medicine MA is still being developed, so the use cases were made with mockups in a test environment. They were not immersed in the real work scenario, with the patient, with the transfusion kit, with the TMT wearing latex gloves, a particular Smartphone to read the appropriate

barcode labels (with a suitable camera, with good Autofocus time, battery life and processing speed) or the hospital's Wi-Fi network.

Future lines

We aim to carry out next studies with qualitative methodologies in user satisfaction after use it in real scenarios. We aim to increase transfusion adverse reactions records; provide greater control at the patient's bedside, improving the effectiveness and efficiency of the transfusion process. We also want to analyze the best strategy, contrasting BYOD vs. the purchase of devices and see which is the minimum hardware requirement when choosing a smartphone [20].

Conclusion

There are multiples devices and applications that represent the workflow for transfusion process using barcode techniques. We design a mobile application applying usability techniques being responsive for smartphones or tablets that use Android OS. We expect this AM to be a tool that helps TMT, achieving greater security in the transfusion process, reducing errors in critical points, improving communication between the different actors involved and facilitating the daily task.

Acknowledgments

This work was carried out thanks to the contribution of a great team of professionals, among them we highlight all the staff of the TMS, especially Dr Walter Scordo, Dr. Camino Pablo, technician Diego Santoro, and Lic. Ortiz Juan who helped and advised us in design and usability of the project.

References

- [1] J. Linden, B. Paul, K. Dressler, A report of 104 transfusion errors in New York State, *Transfusion*. 32 (1992) 601–606. doi:10.1046/j.1537-2995.1992.32792391030.x.
- [2] P.H.B. Bolton-Maggs, H. Cohen, Serious Hazards of Transfusion (SHOT) haemovigilance and progress is improving transfusion safety., *Br. J. Haematol.* 163 (2013) 303–14. doi:10.1111/bjh.12547.
- [3] D. Stainsby, H. Jones, D. Asher, C. Atterbury, A. Boncinelli, L. Brant, C.E. Chapman, K. Davison, R. Gerrard, A. Gray, S. Knowles, E.M. Love, C. Milkins, D.B.L. McClelland, D.R. Norfolk, K. Soldan, C. Taylor, J. Revill, L.M. Williamson, H. Cohen, SHOT Steering Group, Serious Hazards of Transfusion: A Decade of Hemovigilance in the UK, *Transfus. Med. Rev.* 20 (2006) 273–282. doi:10.1016/j.tmr.2006.05.002.
- [4] A. Ohsaka, M. Kobayashi, K. Abe, Causes of failure of a barcode-based pretransfusion check at the bedside: experience in a university hospital, *Transfus. Med.* 18 (2008) 216–222. doi:10.1111/j.1365-3148.2008.00868.x.
- [5] J.C.W. Chan, R.W. Chu, B.W.Y. Young, F. Chan, C.C. Chow, W.C. Pang, C. Chan, S.H. Yeung, P.K. Chow, J. Lau, P.M.K. Leung, Use of an electronic barcode system for patient identification during blood transfusion: 3-Year experience in a regional hospital, *Hong Kong Med. J.* 10 (2004) 166–171.
- [6] C.L. Turner, A.C. Casbard, M.F. Murphy, Barcode technology: its role in increasing the safety of blood transfusion., *Transfusion*. 43 (2003) 1200–9. <http://www.ncbi.nlm.nih.gov/pubmed/12919421> (accessed July 19, 2016).

- [7] S.G. Sandler, A. Langeberg, L. Dohnalek, Bar code technology improves positive patient identification and transfusion safety., *Dev. Biol. (Basel)*. 120 (2005) 19–24.
<http://www.ncbi.nlm.nih.gov/pubmed/16050151> (accessed July 19, 2016).
- [8] A. Porcella, K. Walker, Patient safety with blood products administration using wireless and bar-code technology., *AMIA ... Annu. Symp. Proceedings. AMIA Symp.* 2005 (2005) 614–8.
<http://www.ncbi.nlm.nih.gov/pubmed/16779113> (accessed December 8, 2016).
- [9] K. Miller, C. Akers, G. Magrin, S. Whitehead, A.K. Davis, Piloting the use of 2D barcode and patient safety-software in an Australian tertiary hospital setting, *Vox Sang.* 105 (2013) 159–166.
doi:10.1111/vox.12034.
- [10] J.P. Wallis, Transfusion practice., *Transfus. Med.* 19 (2009) 303.
doi:10.1111/j.1537-2995.2009.02211.x.
- [11] J.H. Nichols, C. Bartholomew, M. Brunton, C. Cintron, S. Elliott, J. McGirr, D. Morsi, S. Scott, J. Seipel, D. Sinha, Reducing medical errors through barcoding at the point of care., *Clin. Leadersh. Manag. Rev.* 18 (n.d.) 328–34.
<http://www.ncbi.nlm.nih.gov/pubmed/15597554> (accessed December 16, 2016).
- [12] Build Amazing Native Apps and Progressive Web Apps with Ionic Framework and Angular, (n.d.). <https://ionicframework.com/> (accessed December 20, 2016).
- [13] AngularJS — Superheroic JavaScript MVW Framework, (n.d.). <https://angularjs.org/> (accessed December 20, 2016).
- [14] Guía para el desarrollo de CDA HL7 Spain_v1.0 Guía para el desarrollo de documentos CDA Subcomité Técnico V3-CDA HL7 Spain, (2007).
- [15] J. Lei, L. Xu, Q. Meng, J. Zhang, Y. Gong, The current status of usability studies of information technologies in China, *Biomed Res. Int.* 2014 (2014). doi:10.1155/2014/568303.
- [16] Usability Testing Basics An Overview, (n.d.). www.techsmith.com (accessed November 16, 2016).
- [17] J. Zhang, M.F. Walji, TURF: toward a unified framework of EHR usability., *J. Biomed. Inform.* 44 (2011) 1056–67.
doi:10.1016/j.jbi.2011.08.005.
- [18] G. Alexander, N. Stagers, A systematic review of the designs of clinical technology: findings and recommendations for future research., *ANS. Adv. Nurs. Sci.* 32 (2009) 252–79.
doi:10.1097/ANS.0b013e3181b0d737.
- [19] N. Stagers, D. Kobus, Comparing response time, errors, and satisfaction between text-based and graphical user interfaces during nursing order tasks, *J. Am. Med. Inform. Assoc.* 7 (2000) 164–76.
<http://www.ncbi.nlm.nih.gov/pubmed/10730600> (accessed December 18, 2016).
- [20] J. Bradley, BYOD: una perspectiva global, *Ieee.* (2012) 21.
https://www.cisco.com/web/about/ac79/docs/re/byod/BYOD_Horizons-Global_LAS.pdf.

Address for correspondence

MD Alborno Mario Alejandro

mario.alborno@hospitalitaliano.org.ar

Assessing Provider-Generated Free-Text Quality in EHR-Integrated Handoff Notes

Elliot G Arsoniadis^{a,b}, Steven J Skube^a, Treva M Bjerke^c, Bryan Jarabek^c, Genevieve B Melton^{a,b}

^a Department of Surgery, University of Minnesota, Minneapolis, MN, USA

^b Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

^c Fairview Health Services, Minneapolis, MN, USA

Abstract

Handoff notes are increasingly integrated within electronic health record (EHR) systems and often contain data automatically generated from the EHR and free-text narratives. We examined the quality of data entered by providers in the free-text portion of our institutional EHR handoff tool. Overall, 65% of handoff notes contained at least one error (average 1.7 errors per note). Most errors were omissions in information around patient plan/management or assessment/diagnosis rather than entry of false data. Factors associated with increased error rate were increasing hospital day number; weekend note; medical (vs. surgical) service team; and authorship by a medical student, first or fourth year resident physician, or attending physician. Our findings suggest that errors are common in handoff notes, and while these errors are not completely false data, they may provide individuals caring for patients an inaccurate understanding of patient status.

Keywords:

Patient Handoff; Patient Transfer; Electronic Health Records;

Introduction

Within patient care, a handoff refers to the process by which clinicians transfer the care of patients from one clinician to another. This process occurs with the transfer of patients between clinical settings, such as from the Emergency Department to the inpatient unit, or when patients transfer between different care teams without necessarily changing physical location, such as when daytime care teams change to nighttime “on-call” teams. This latter type of care transfer between the primary daytime team to the “on-call” team is recognized to be an important source of preventable medical errors [1].

Handoff notes are cognitive aids that are created in order to aid in the transfer of patient care from one clinician (or team of clinicians) to another [2]. Increasingly, handoff notes are being integrated within the Electronic Health Record (EHR), rather than existing as stand-alone documents outside of the EHR [2-4]. EHR-integrated handoff notes may be completely populated with data automatically generated from the EHR, but more often contain both automatically generated data as well as free-text narrative data authored by members of the care team [4]. Patient summaries and management plans are two of the most important items within handoff notes, and almost always require direct free-text entry rather than automatic generation from data in other parts of the EHR [5].

The goal of this study was to assess the quality of free-text narrative data in handoff notes generated by physicians within an institutional EHR in an academic inpatient setting. First, we sought to learn which providers and medical teams at our tertiary-care institution (composed of medical students, resident physicians in training, attending physicians, and Advanced Practice Providers (APP)) were authoring handoff notes. Then, we evaluated the quality of data in these free-text narratives by analyzing their accuracy and completeness using daily progress notes and other data within the electronic chart as the gold standard. Finally, we attempted to identify patterns in the frequency and types of errors encountered, and to ascertain factors associated with errors in free-text data.

Methods

Introduction of an EHR-integrated Handoff Tool

In June 2016, our institution introduced an electronic handoff tool incorporated within our institutional EHR (Epic Systems, Verona, Wisconsin, USA). This tool was optimized and locally adapted to be more user friendly and legible by a team of eight physician informaticists and two information technology builders. The design was based on prior experience with the handoff tool at several other peer academic institutions and work with several inpatient-based provider groups. The handoff tool template included two free-text text boxes, labeled “Patient Summary” and “To-Do”. Authors could enter free-text within either box. The handoff tool could be accessed and viewed within the EHR by any member of the care team, including medical students, resident physicians, APPs, and attending physicians. A print option was also available, which included any free-text entered in either the “Patient Summary” or “To-Do” boxes as well as automatically generated patient information, including patient demographics, vital signs, and laboratory data.

Obtaining Physician-generated Handoff Notes

Free-text data was collected from handoff notes each evening from 7-9PM during a six week period. Patients included were on both surgical and medical service teams in general inpatient units. Patients in critical care settings were not included. Daily handoff note collection began on the day of admission and continued each day until the day of discharge. In addition to the content of the handoff “Patient Summary” and “To-Do” free-text text boxes, we also collected data on the date and time of data entry, the author’s specialty and level of training, as well as whether the note occurred on a weekday or weekend.

The free-text information in the handoff note was then compared to the information in the patient chart, including daily progress notes, laboratory data, imaging studies, and orders, to assess for accuracy. In addition to assessing the accuracy of the free-text information written by handoff note authors, physician-raters also assessed for missing information omitted from the handoff note.

Prior to initiation of data collection, two physician-raters met and formulated an initial schema to define quality issues in handoff notes. Missing information was defined as any data missing from key portions of the patient's diagnosis and subsequent clinical course or key elements of the management plan. History and Physical Exam notes written on the day of admission and subsequent daily progress notes were used as the standard against which to compare the free-text data in the handoff note. In the initial schema, data included in the Assessment and Plan portion of the admission History and Physical Exam note or daily progress notes that was omitted from either the free-text "Patient Summary" or "To-Do" portions of the handoff note were deemed missing information. Missing significant results of imaging studies or laboratory tests, as well as notes from consulting medical teams were also defined as missing information. Again, the daily progress note and subsequent data generated throughout the clinical work day constitute basic information that should be included in the free-text portions of the handoff note. After this schema was defined, physician-raters separately reviewed thirteen daily handoff notes and compared their individual assessments. Differences in assessment were discussed and agreed upon, and the final schema was adjusted to account for the variation in physician-rater assessment.

After six weeks, the physician-raters had collected free-text information from 368 handoff notes. The quality of these notes was assessed using the aforementioned schema and errors were recorded. We then sought to describe the nature of the errors uncovered from free-text data in handoff notes, using a schema originally described by Arora and colleagues for assessing medication-related errors in handoff notes [6]. We adjusted the schema to fit our broader goals of defining both medication and non-medication related errors in handoff notes. Errors were initially classified as either errors of commission or errors of omission. Errors of commission were defined as those errors committed by authors where incorrect information was entered into the handoff note. For instance, an author may have entered "patient receiving ciprofloxacin" but on review of the active orders in the patient chart, the patient may have been on an antibiotic other than ciprofloxacin. Entered information that was no longer relevant was also classified as an error of commission. For instance, if the handoff note states "patient to receive chest CT if respiratory status declines" and review of imaging studies show that a chest CT had already been completed that day, this would also be deemed an error of commission. Errors of omission were those where a piece of clinically relevant information related to the diagnosis/clinical course or management plan was omitted from the free-text data in the handoff note (Table 1).

In addition to assigning errors as either those of commission or omission, errors were also grouped according to whether they were an error in assessment/diagnosis or an error in plan/management. Errors were assigned this label based on whether the information would more likely be included in the Assessment or the Plan portion of a traditional SOAP (Subjective, Objective, Assessment, Plan)-style daily progress note. For instance, errors regarding surgical procedures that

had occurred during hospital admission or important clinical events, such as the development of Acute Kidney Injury, were errors of assessment/diagnosis. Errors such as omitting that Gastroenterology was consulted or listing the wrong antibiotic for treatment of pneumonia were considered errors in management/plan.

Finally, errors were assigned as either "New" or "Continued". Errors appearing for the first time in handoff notes for a hospitalization were defined as "New", whereas the same error included in a patient's subsequent handoff notes was considered "Continued". After the number and types of errors were determined for each handoff note, the overall percentage of notes with errors were compared across disciplines, levels of training, hospital day, as well as if the note was written on a weekday or weekend.

Table 1 – Examples of Handoff Note Errors

Error	Data in Handoff Note	Data in EHR (Gold Standard)
Commission in Assessment/Diagnosis	Patient is post-operative day 0	Patient is post-operative day 2
Commission in Management/Plan	Patient is on ciprofloxacin	Patient is on piperacillin
Omission in Assessment/Diagnosis	No mention of deep vein thrombosis (DVT)	DVT diagnosed on day 4
Omission in Management/Plan	No mention of heparin drip	Heparin drip started for DVT on day 4

Table 2 – Handoff Note Author Demographics and Other Variables

Variable	Number (%Total)
Author	
Medical Student	28 (8%)
PGY-1	188 (51%)
PGY-2	99 (27%)
PGY-3	19 (5%)
PGY-4	10 (3%)
Attending physician	17 (4%)
APP	8 (2%)
Day of Week	
Weekday	269 (73%)
Weekend	99 (27%)
Clinical Service Team	
Medical	181 (49%)
Surgical	187 (51%)
Total	368 (100%)

Results

Handoff Note Demographics

Overall, 368 handoff notes were collected and evaluated during the study period. Handoff note authors ranged in level of training from medical students, to resident physicians in post-graduate years (PGY) 1 through 4, APPs (Nurse Practitioners and Physician Assistants), and attending physicians. The majority of handoff notes were written by resident physicians in PGY-1 (51%) and PGY-2 (27%) (Table

2). There was a nearly even distribution in handoff notes written on patients on internal medicine (49%) and surgery service teams (51%). Surgery teams included General, Bariatric, Colorectal, Thoracic, Transplant, Gynecology, and Ear Nose and Throat (ENT) surgery. The majority of the handoff notes were written on Hospital Day (HD) 1 (27%) and HD2 (21%). Among those patients included in our study, length of stay ranged from 1-25 days.

Handoff Errors

Overall, 635 errors were discovered in 368 handoff notes, an average of 1.7 errors per handoff note. In 65% of handoff notes, at least one error was found. The vast majority of errors were rated as either plan/management omission (54%) or assessment/diagnosis omission (32%). There were far fewer errors of commission in both plan/management (12%) or assessment/diagnosis (2%). Of the omissions in plan/management, nearly half (48%) were rated as “New” errors. Only 32% of omissions in assessment/diagnosis were rated as “New” errors, with the majority (68%) being rated as “Continued” errors. Over half of the errors of commission in plan/management (59%) and assessment/diagnosis (60%) were rated as “New” errors (Table 3).

Table 3 – Errors in Handoff Notes

Type of Error	Number (%Total)	New Errors (%)
Omission in Plan/Management	343 (54%)	165 (48%)
Omission in Assessment/Diagnosis	201 (32%)	65 (32%)
Commission in Plan/Management	76 (12%)	45 (59%)
Commission in Assessment/Diagnosis	15 (2%)	9 (60%)
Total	635 (100%)	284 (45%)

Error Rate and Type by Hospital Day

The rate of having at least one error detected in a handoff note and the average number of errors per note increased with each subsequent day in the hospital. Forty-seven percent of handoff notes written on HD1 had at least one error detected with an average of 0.9 errors per note. This increased to 57% on HD2, with 1.19 average number of errors per note. A subsequent increase was noted on HD3, 4, and 5 as well, where 65%, 74%, and 78% of notes, respectively, had at least one error detected (Table 4). Average number of errors per note also increased, with 1.69, 1.91, and 2.26 errors respectively. Handoff notes written on HD6 or later had the highest rate of errors (86%), and the highest average number of errors per note (2.95) (Table 4).

The types of errors also changed with increasing hospital day. On HD1, 61% of errors were omissions in plan/management and 22% were related to omissions in assessment/diagnosis. Only 6% were errors of commission in plan/management and 2% errors of commission in assessment/diagnosis. This distribution of types of errors stayed relatively consistent for HD2-5 (Table 5). However, for handoff notes written on HD6 and beyond, omissions in plan/management decreased (42%)

of total errors) while omissions in assessment/diagnosis increased (44% of total errors) (Table 5).

The percentage of errors that were rated as “New” versus “Continued” also changed with increasing hospital day. On HD2, 76% of omissions in plan/management and 55% of omissions in assessment/diagnosis were rated as “New.” By HD6 and beyond, only 17% of omissions in plan/management and 13% of omissions in assessment/diagnosis were rated as “New” (Table 5).

Table 4 – Errors by Hospital Day, Day of Week, Clinical Service and Provider Level

Variable	Total Number of Errors	Percentage of Notes with Error	Average Number of Errors per Note
Hospital Day			
1	88	47%	0.90
2	92	57%	1.19
3	91	65%	1.69
4	67	74%	1.91
5	52	78%	2.26
6+	245	86%	2.95
Day of Week			
Weekday	457	62%	1.70
Weekend	245	86%	2.95
Clinical Service			
Medical	374	72%	2.07
Surgical	259	55%	1.39
Provider			
Medical Student	62	85%	2.38
PGY-1	310	68%	1.65
PGY-2	137	54%	1.38
PGY-3	19	47%	1.0
PGY-4	14	70%	1.40
APP	7	38%	0.88
Attending Physician	72	82%	4.24
Overall	368	65%	1.7

Error Rate and Type by Weekday versus Weekend

Handoff notes written on a weekend had a higher rate of having at least one error detected (73%) and higher average error (1.8) versus those notes composed on a weekday (62% error rate and 1.7 average errors) (Table 4).

Omissions in plan/management accounted for 56% of errors in handoff notes written on weekdays, and 48% of errors in notes written on weekends. Omissions in assessment/diagnosis accounted for 32% of errors in both weekday handoff notes and weekend handoff notes. Sixteen percent of errors in handoff notes were related to errors of commission in plan/management on weekends, versus only 11% of errors on weekdays.

Error Rate and Type by Service Team

The percentage of notes with at least one error detected and average number of errors detected per handoff note varied by service team. Fifty-five percent of handoff notes written by authors on surgical service teams had at least one error detected, and an average of 1.39 errors per note. In contrast, 72% of handoff notes written by authors on medical service

teams had at least one error detected, with an average of 2.07 errors per note (Table 4).

Sixty-six percent of errors in handoff notes on medical service teams were omissions in plan/management, while only 37% of errors on surgical service teams were omissions in plan/management. Conversely, on surgical service teams 44% of errors were omissions in assessment/diagnosis while on medical service teams omissions in assessment/diagnosis accounted for only 23% of errors.

Table 5 – Distribution of Error Types by Hospital Day

HD	Plan/Management Omission (%New)	Assessment/Diagnosis Omission (%New)	Plan/Management Commission (%New)	Assessment/Diagnosis Commission (%New)
1	61% (100%)	22% (100%)	6% (100%)	2% (100%)
2	59% (76%)	22% (55%)	17% (94%)	2% (100%)
3	56% (35%)	27% (56%)	13% (67%)	3% (67%)
4	64% (42%)	24% (25%)	7% (75%)	4% (33%)
5	62% (38%)	23% (17%)	15% (50%)	0
6+	42% (17%)	44% (13%)	12% (31%)	2% (40%)

Error Rate and Type by Training Level of Author

The majority of handoff notes were written by resident physicians in PGY-1 or PGY-2. Sixty-eight percent of handoff notes written by PGY-1 physicians had at least one error detected, just slightly above the overall rate of 65% for all training levels. Fifty-four percent of handoff notes written by PGY-2 physicians had at least one error detected. PGY-1 physicians had an average of 1.65 errors per handoff note and PGY-2 physicians had an average of 1.38 errors per note. Overall, there were an average of 1.73 errors per handoff note for all authors (Table 4).

Nearly half of errors in handoff notes (49%) written by PGY-1 physicians were omissions in plan/management, while well over half (60%) of errors in notes written by PGY-2 physicians were omissions in plan/management. Thirty-three percent of errors were omissions in assessment/diagnosis for PGY-1 physician-generated notes, and 20% of errors were omissions in assessment/diagnosis for PGY-2 physician-generated notes.

Only 8% of handoff notes were written by medical students, with an average of 2.38 errors per note. Of these notes, 85% had at least one error detected. PGY-3 and PGY-4 physicians authored 5% and 2.7% of notes, respectively. Forty-seven percent of handoff notes authored by PGY-3 physicians and 70% of notes authored by PGY-4 physicians had at least one error detected.

Few notes were authored by attending physicians or APPs. Eighty-two percent of handoff notes authored by attending physicians had at least one error detected, with an average of 4.24 errors per note. Thirty-eight percent of handoff notes written by APPs had at least one error detected, with an average of 0.88 errors per note.

Discussion

In the present study, we analyzed free-text data within a series of handoff notes written using an EHR-integrated handoff tool. The majority of authors were PGY-1 or PGY-2 physicians. Over half of the errors encountered were omissions in plan/management and nearly one third were omissions in assessment/diagnosis, reflecting a lack of completeness, rather than gross inaccuracies, as the major source of error. When examining the errors of commission in plan/management that had been entered into handoff notes, 63% (47 out of 75) were due to failure to update a previously accurate plan that had since changed. Again, this shows that a major contributor to false information was the failure to update previously true information, rather than the direct entry of erroneous data. This points to a deficiency in effort on the part of handoff note authors, rather than a deficiency in knowledge, as a major cause of inaccurate and incomplete data in handoff notes.

The Role of Hospital Day and Information Decay

A major contributor to the amount of errors detected in handoff notes was the length of hospital stay, with increasing days correlating with increasing errors. Forty-seven percent of handoff notes had at least one error detected on HD1. By HD6 or beyond, 86% of handoff notes had at least one error detected. Increased length of hospital stay correlated with an increased risk of errors as new clinical information is gained, along with a higher probability of propagating previous errors. Arora et al noticed this in their study of medication discrepancies, noting that 63% of errors persisted beyond their index case (6). This corresponds to our data, where 24% of omissions in plan/management and 45% of omissions in assessment/diagnosis were rated as “Continued” on HD2, yet by HD6 and beyond, 83% of omissions in plan/management and 87% of omissions in assessment/diagnosis were rated as “Continued.” Interestingly, the majority of errors initially were omissions in plan/management, but by HD6 and beyond, there were nearly equal omissions in assessment/diagnosis. This reflects the fact that while authors were relatively accurate at recording initial diagnoses and clinical events, accuracy waned as hospital day, and presumably clinical complexity, increased.

Weekday versus Weekend Handoff Notes

Handoff notes from weekends were more likely to have at least one error (73% vs 62%) and had a higher average number of errors per note (1.80 vs 1.70) compared to notes written on weekdays. Other clinical phenomena have been associated with weekend care. Admissions and surgical interventions on weekends have previously been associated with higher mortality [7], length of stay [8] and hospital-acquired conditions [9]. These associations could be due to decreased numbers of physicians on weekend care teams. With fewer physicians present to aid in clinical work, those physicians who are present may prioritize other clinical activities at the expense of updating handoff notes.

The Role of Clinical Service Team and Author Training Level

To our knowledge, this is the first study to compare rates of errors in handoff notes by specialty. Handoff notes for patients on medical service teams were more likely to contain at least one error (72% vs 55%) and had a higher average number of errors per note (2.07 vs 1.39) compared to surgical

service teams. One possible explanation is the level of training of the authors. All attending physicians, who had the highest rates of errors, were on medical service teams while the vast majority of APPs, who had the lowest rates of errors, were on surgical service teams. Notably, medical service teams were more likely to omit items from plan/management while surgical teams were more likely to omit items from assessment/diagnosis, possibly reflecting differences in handoff data prioritization between specialties.

The percent of handoff notes with at least one error detected, as well as average number of errors per note, varied by author level of training. Increased training initially correlated with improvements in error rate, with PGY-2 and PGY-3 physicians having fewer errors than medical students or PGY-1 physicians. The benefits of increased training, knowledge, and clinical acumen could account partially for the improvement in handoff note errors. However, increased training beyond PGY-3 (PGY-4 and attending physicians) was associated with a higher error rate. This could reflect that more experienced physicians (beyond PGY-3) rely less on the handoff note as a comprehensive summary of the patient's clinical course and management plan and thus require a less granular version to serve its purpose as a cognitive aid [10]. This also suggests that physicians-in-training interface with the EHR in a qualitatively different way as training progresses and attending-level status is reached [11].

Study Limitations

While the handoff note represents an important cognitive artifact, there are other verbal aspects of the handoff process that we could not evaluate in this study. Further studies might include a focus on both the verbal and written aspects of handoff. Studying handoff notes at different timepoints throughout the year may also show a difference in error rates, especially at teaching institutions where physicians-in-training progress in their clinical acumen throughout the year.

Finally, this study utilized a schema we devised to determine what information was important for inclusion in the handoff note. Decisions about the accuracy and completeness of handoff notes were based on information recorded in the remainder of the electronic chart. Further work might include a larger body of physician-raters from various specialties and training levels to further validate the present methods.

Conclusions

Increasingly, handoff notes are being integrated within the EHR. We found free-text data in EHR-integrated handoff notes frequently contain errors. The majority of these errors are related to the omission of information regarding both plan/management and assessment/diagnosis. Far fewer errors are directly related to the entry of erroneous information. Further work should focus on efforts to increase the ease with which accurate handoff notes can be generated. This will rely on efforts from clinicians, members of the health IT community, and experts in human factors to ease the process of generating high-quality handoff notes, rather than simply increasing the amount of automatically generated data in the handoff note. It will also involve improvements in clinician training, emphasizing the importance of high-quality handoff notes and recognizing the relevance of these notes in providing high-quality patient care.

Acknowledgments

The authors would like to thank the University of Minnesota Department of Surgery, Fairview Health Services, and University of Minnesota Physicians/MHealth for their support of this work.

References

- [1] V. Arora, J. Johnson, D. Lovinger, H.J. Humphrey, D.O. Meltzer. Communication failures in patient sign-out and suggestions for improvement: a critical incident analysis. *Qual Saf Health Care* **14** (2005), 401-407.
- [2] A.J. Starmer, T.C. Sectish, D.W. Simon, C. Keohane, M.E. McSweeney, E.Y. Chung, et al. Rates of medical errors and preventable adverse events among hospitalized children following implementation of a resident handoff bundle. *JAMA* **310** (2013), 2262-2270.
- [3] J.P. Palma, P.J. Sharek, C.A. Longhurst. Impact of electronic medical record integration of a handoff tool on sign-out in a newborn intensive care unit. *J Perinatol* **31** (2011), 311-317.
- [4] J. Abraham, T. Kannampallil, V.L. Patel. A systematic review of the literature on the evaluation of handoff tools: implications for research and practice. *J Am Med Inform Assoc* **21** (2014), 154-162.
- [5] G. Rosenbluth, J.F. Bale, A.J. Starmer, N.D. Spector, R. Srivastava, D.C. West, et al. Variation in printed handoff documents: results and recommendations from a multicenter needs assessment. *J Hosp Med* **10** (2015), 517-524.
- [6] V. Arora, J. Kao, D. Lovinger, S.C. Seiden, D. Meltzer. Medication discrepancies in resident sign-outs and their potential to harm. *J Gen Intern Med* **22** (2007), 1751-1755.
- [7] R. Ricciardi, P.L. Roberts, T.E. Read, N.N. Baxter, P.W. Marcello, D.J. Schoetz. Mortality rate after nonelective hospital admission. *Arch Surg* **146** (2011), 545-551.
- [8] J.P. Rothman, J. Burcharth, H.C. Pommergaard, J. Rosenberg. Cholecystectomy During the Weekend Increases Patients' Length of Hospital Stay. *World J Surg* **40** (2016), 849-855.
- [9] F.J. Attenello, T. Wen, S.Y. Cen, A. Ng, N. Kim-Tenser, N. Sanossian, et al. Incidence of "never events" among weekend admissions versus weekday admissions to US hospitals: national analysis. *BMJ* **350** (2015), h1460.
- [10] E.S. Patterson, E.M. Roth, D.D. Woods, R. Chow, J.O. Gomes. Handoff strategies in settings with high consequences for failure: lessons for health care operations. *Int J Qual Health Care* **16** (2004), 125-132.
- [11] G. Hultman, J. Marquard, E. Arsoniadis, P. Mink, R. Rizvi, T. Ramer, et al. Usability testing of two ambulatory EHR navigators. *Appl Clin Inform* **7** (2016), 502-515.

Address for correspondence

Genevieve B. Melton, MD, PhD
 Institute for Health Informatics and Department of Surgery
 420 Delaware Street SE, Mayo Mail Code 450
 Minneapolis, MN 55455
 Tel: +1 612-625-7992
 E-mail: gmelton@umn.edu

Accessing Reliable Health Information on the Web: A Review of the HON Approach

Célia Boyer^a, Arnaud Gaudinat^d, Allan Hanbury^e, Ron D. Appel^{b, c}, Marion J. Ball^{b, c}, Michel Carpentier^{b, c}, Jan H. van Bommel^{b, c}, Jean-Paul Bergmans^b, Denis Hochstrasser^c, Donald Lindberg^c, Randolph Miller^b, Jean-Claude Peterschmitt^{b, c}, Charles Safran^b, Michèle Thonnet^b, Antoine Geissbühler^b

^a Executive Director – Health On The Net Foundation, Geneva, Switzerland

^b Foundation Council members – Health On The Net Foundation

^c Co-founders – Health On The Net Foundation

^d BiTeM group, HES-SO/HEG Geneva, Information Science Department, Carouge, Switzerland

^e Institute of Software Technology & Interactive Systems, TU Wien, Vienna, Austria

Abstract

Accessing online health content of high quality and reliability presents challenges. Laypersons cannot easily differentiate trustworthy content from misinformed or manipulated content. This article describes complementary approaches for members of the general public and health professionals to find trustworthy content with as little bias as possible. These include the Khresmoi health search engine (K4E), the Health On the Net Code of Conduct (HONcode) and health trust indicator Web browser extensions.

Keywords:

Quality Indicators; Internet; Trusted Health Information

Introduction

The Web has become an important and significant health-related resource. (N.B., the terms resource, website and webpage are used interchangeably herein for health content on the Web.) By providing copious medical and healthcare information, the Web has empowered the public. Some 60% of Europeans [1] and 72% of Americans [2] have used the Web to address healthcare questions. Internet use offers many conveniences, including accessibility 24 hours a day, anonymity when conversing with others, and analyses on a wide range of subjects. The Web provides information on how to contact local and national experts, access to sage opinions (e.g., best therapies, effectiveness research), and connectivity to a massive quantity of health information resources. Alarming, some 75% of people using the Web for health purposes do not differentiate among the facts they obtain in terms of accuracy or credibility [3].

A 2013 Pew Research survey indicated that eight out of ten online health inquiries start by using a search engine [2]. Somewhat uniquely, Web-based health information can directly impact a person's health status, providing benefits, but also harm - in a few extreme cases, resulting in death [4]. Naive users do not realize that search engine results mix trustworthy information with unreliable and even purposefully manipulated health information. The most commonly used search engines currently provide no support to differentiate quality assessments of health content. Even worse, general search engine results can lead to biased medical content (deviation from the truth"), and can lead users to make inappropriate

healthcare decisions [5]. PageRank was identified early as a promising way for health websites to indicate to consumers that they were providing quality information to consumers [6].

However, mechanisms underlying page rankings, e.g., hyperlinks and browsing history, merely indicate popularity of a webpage is [7]. Many organizations have previously attempted to guide Web users to high-quality health websites, but this remains a challenge. For example, Google investigated possible statistical estimating schemes to judge the correctness of

facts [7]. Nevertheless, in the rapidly evolving health domain, facts are not absolute – the best test or therapy for a condition today may not be so tomorrow. Information correct for one person's situation may not be for others. Thus a crucial need exists to assess the trustworthiness of a given health Web resource. This paper presents recent new quasi-automated methods to filter healthcare web site content to assess trustworthiness and readability. The ultimate goal is to promote laypersons' easy access to quality health information [8, 9].

The problem of trust

The concept of trust is elusive to define because a multitude of factors contribute to it. Grandison and Sloman [8] defined the quality of a page as content targeted at the right audience. Gil and Artz [9] listed 19 different factors affecting how users determine trust in webpage content, including: topic, context and criticality, popularity, authority, recommendation, bias, appearance, honesty and currency of information.

In 1996, the Health On the Net (HON) Foundation determined that the International Committee of Medical Journal Editors (ICMJE – www.icmje.org/) conventions and recommendations for printed medical journals could be applied to online information. Thus, HON developed the HONcode, a set of ethical, honesty, transparency and quality standards related to health website content production. Note that HONcode certification provides a metric that determines if the processes underlying a website's construction and maintenance conform to standards of excellence; it does not evaluate the veracity of the site's content per se. The HONcode certification process implies that the health website editors are both motivated and committed as they need to invest time to meet HONcode criteria into the future. The website editors receive no direct financial return or incentive for such efforts from HON. Moreover,

in 2014 the annual assessment of certified websites was changed into a contribution-based program enabling HON to continue offering certification. Prices range from 50 euros for not-for-profit websites to 325 euros for high-ranked commercial websites. Certified health websites agree to display the HONcode seal on the website (Figure 1), to be continuously scrutinized and to implement the recommendations made by HON.



Figure 1 – HONcode seal displayed on certified websites

Based on more than two decades of research and pragmatic experience accrediting websites using HONcode criteria [10], the authors now believe that those criteria capably capture the trustworthiness of a health Web resource. Studies have demonstrated that websites that conform to HONcode quality standards contain more reliable health information than randomly selected health-related websites [18]. The presence of the HONcode symbol on a website informs the user that the site respects a quality standard. Thus the user, when faced with multiple, contradictory and sometimes questionable information, can trust those sites that are HONcode certified. Unfortunately, HON lacks the resources to review and revisit all health-related websites continuously. Also, HONcode certification is a voluntary process whereby health website editors must request that HON reviews their sites, which requires awareness of the HONcode initiative. So, the problem at hand is how to estimate the trust level of an arbitrary health website by quasi-automated means.

Previous relevant work on health website trustworthiness

Other groups have explored natural language processing (NLP) approaches to facilitate access to quality health information [11, 12]. The authors and colleagues have also examined automated identification of information trustworthiness using the HONcode criteria combined with an experimental multilingual automated detection system [13, 14]. The latter NLP-based approach also included readability-level scoring, which rates how easy it is for the average user to understand the web page content. In addition to developing methods that analyze content that users have found, HON has also developed services that help Web users to directly access trustworthy health information. HON has actively participated in the development of the Khresmoi for Everyone (K4E – <http://everyone.khresmoi.eu>) health search engine, which provides access to trustworthy health websites. The JAMA benchmark criteria, the DICTER score (<http://www.discrim.org.uk>) and Medlineplus.gov and the HONcode have been widely covered and compared in the literature for assessing health and medical-related websites [15, 16, 17]. The Table 1 below summarizes the specificities of the four instruments.

Why automate HONcode certification?

Previous approaches to the HON certification process have been described elsewhere [10, 13]. Because HONcode certification has been until now carried out manually, the number of sites that can be assessed by HON reviewers on a daily basis is limited. This article focuses on quasi-automated detection of compliance with the HONcode criteria in a manner that complements the human task. Yet, ample room exists for greater awareness of online health information quality amongst the

general public. Additionally, it is not easy to determine which websites are certified when using a general search engine. This is why, in addition to the HONcode Web browser extension, the automated detection of HONcode principles has been studied and developed.

Table 1 – Comparison of main initiatives to identify trustworthy online content

Initiatives	Types of initiative	Main differences
DICTER	16 questions	No implementation
JAMA benchmark criteria	4 main criteria	No implementation
HONcode certification	8 criteria Certification process conducted by trained HON health professionals	See Table 2 - Voluntary approach - Motivate health website editors to improve the production process of their sites - Search engines with access to certified websites - Manually assessed and curated according to published guidelines
MedlinePlus	Manually selected webpages	- Manually curated list according to 15 criteria - From the U.S.A. government

Methods

HONcode quality criteria

The proposed approach is based on the HONcode. Table 2 lists the HONcode principles.

Table 2 – The eight HONcode principles

Principle	Detail
HC1 - Authority	Indicates the qualifications of the authors
HC2 - Complementarity	Information supports, does not replace the doctor-patient relationship
HC3 - Privacy policy	Respects the privacy and confidentiality of personal data submitted to the site by the visitor
HC4 - Attribution of reference criteria date	Cites the source(s) of published information Dates medical and health pages
HC5 - Justifiability	Backs up claims relating to benefits and performance
HC6 - Transparency	Presentation is accessible; contact information is present Identifies funding sources
HC7 - Financial disclosure	Identifies funding sources
HC8 - Advertising policy	Clearly distinguishes advertising from editorial content

Automated detection of HONcode principles

HON has conducted extended research on benchmarking and assessing natural language processing (NLP) methods for multilingual automated detection of HONcode principles [13, 14, 19]. The first step of the manual and any automated certification process is to determine where on the candidate web site information relevant to each HONcode criterion appears. This requires around 25% of the time of manual HON assessors. The automated machine learning algorithm training data set (ground truth) comprises the notes that previous expert manual HON reviewers created during their reviews – and include extracts of texts justifying that each principle was met [13, 14] in real life settings. For each HONcode principle, a specific classifier has been created, except for the Attribution principle (HC4), which has been divided into one classifier focusing on references (HC4-Reference) and one on date (HC4-Date), giving nine distinct classifiers (Figure 2). The resulting panel of NLP algorithms has been tested, compared, evaluated, fine-tuned and applied in order to develop classifiers for each of

the HONcode principles. This resulted in the selection of the Naïve Bayes (NB) supervised machine learning algorithm [13,14]. The classifiers are used to identify the presence of the HONcode criteria into health webpages in English and French.

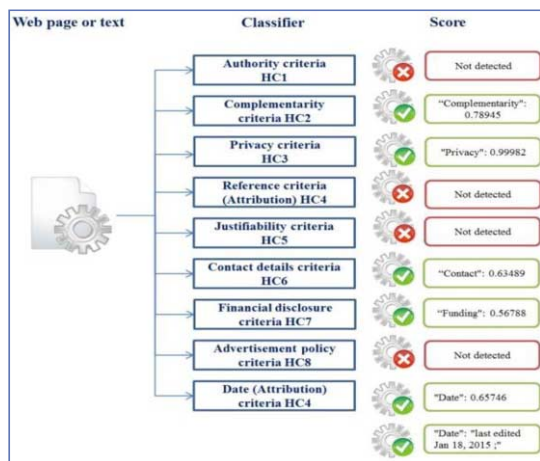


Figure 2– 9 classifiers for automated detection of HONcode criteria

All of the classifiers are applied to Web page as a page can mention several principles. All the principles detected within a given domain name are associated to that domain name. A page is considered classified as per a given principle once it has reached a certain threshold set by the benchmarks [13, 14]. The classification is done using the key terms specific to each HONcode principle as determined by machine learning. The text in colored boxes in Figure 3 illustrates how the classifier score is elaborated based on a collection of key terms identified. The HONcode reviewer extracts the full text corresponding to the principle, as shown in yellow in Figure 3. However, the automated detection of the presence of HONcode criteria only provides an overall estimate of a health website's level of credibility.

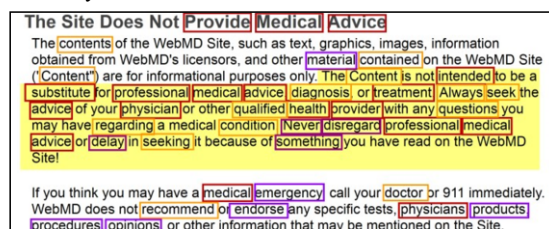


Figure 3– Terms detected by the automated system vs. those detected manually by HON assessor for HC2-Complementarity criterion

An evaluation of how the classifiers work on more complex and real health websites, as opposed to simply extracts, as was done for the systematic evaluation of the classifiers, was conducted [13]. This evaluation comprised 27 highly complex English-language health websites evaluated both by the expert HONcode reviewers and the automated HONcode system. Prior to the use of the classifiers on the health websites, various processes to extract the content of the websites were run. These websites were then parsed by the automated system. In parallel, the websites were assessed manually by the HONcode experts. The average manual HONcode evaluation time for a health website is 80 minutes. Human reproducibility of

the results of a HONcode certification ranges from 80% to 95% depending on the principles [13]. Lastly, the results are compared with the HONcode experts verifying the compliance of HONcode criteria (the automated system only detects the presence of words characterizing the HONcode criteria). Globally the automated detection of HONcode criteria performs well for most of the criteria. This evaluation has led to improvements in the criteria related to HC2-Complementarity, HC4-Date and HC8-Advertising policy. Further investigation shows that for the HC4-Date criterion, the named entity recognition (NER) technique was better adapted than machine learning and a sliding window as a classification unit was necessary for the HC2-Complementarity criterion in order to detect text with other criteria included in the same page. Upon further evaluation, these improvements demonstrate their efficiency [19] and thus have been adopted and used within the automated detection of HONcode criteria (Figure 5).

Readability level of health content

The goal of introducing a readability level is to determine how difficult it is to understand medical or health webpages. The idea is to provide users with access to documents targeted at their level of understanding, a level that evolves over time. How difficult is it to understand the health information on a given webpage? Gil and Artz [9] define access to information adapted to the audience as an indicator of quality. Readability levels have been widely investigated from linguistic and stylistic points of view [20], but little has been done at the level of the medical and health domain where complex terminologies are often used, rendering information difficult to understand.

Further investigations have been conducted using machine learning algorithms to categorize the complexity of health information. The readability level is calculated taking into consideration the length of a sentence and the vocabulary and syntax within the medical context [14]. A readability level in one of three categories—easy, moderate or difficult—is then assigned to each health webpage, text or document analyzed.

Results

HON proposes two ways for the public at large to access quality health information: (1) through a dedicated health search engine (such as K4E) with a selection of trustworthy and adapted resources available for the readers; (2) through a browser functionality, including the filtering and highlighting of certified HONcode websites, the automated detection of HONcode criteria and a readability indicator. As general search engines are most often the gateway that laypeople use to access health information online [21], the latter has been an important strategy. In addition, the automated system to detect HONcode criteria aims to assist the HON assessors within the evaluation of a health website's HONcode conformity. Therefore, the automated assistance in conducting HONcode reviews may help in accelerating the current time-consuming tasks of HONcode certification and ongoing surveillance.

Dedicated health search engine

The K4E search engine offers a curated list of online health resources that have been manually checked for quality. K4E includes various functionalities related to quality, such as query formulation, readability and trust indicators. K4E is an alternative to general search engines as it reduces content biases in the search results as the index is curated, the results are based only on the relevance according to search terms, and the

organizations behind the search engine are impartial (no advertising). The automated HONcode detection system and the readability level have been integrated into K4E as a demonstrator (Figure 4): a) the trust level indicates to what extent the user can rely on the information provided on the site as determined by the automated HONcode detection system; and b) the readability level indicates how hard it is to understand the health information. The results in a) provide the overall percentage of the HONcode criteria automatically detected and lists those not found for a given health domain.

Usability testing was conducted in a —real-life! setting with members of the general public seeking health information online. The evaluation included two types of feedback: the informal feedback given during the session and recorded by the usability testing software Morae (TechSmith's usability testing software, 2014, version 3.3.3.) and the answers to the standard system usability scale questionnaire (SUS), which presented 10 standard SUS statements used to measure usability perceptions and 17 specific statements related to the K4E search engine. The usability test confirmed that the level of —readability and trust indicator score are important user requirements, with a score of 4.34 out of 5 in the Likert scale: strongly disagree (1) to strongly agree (5) [22].

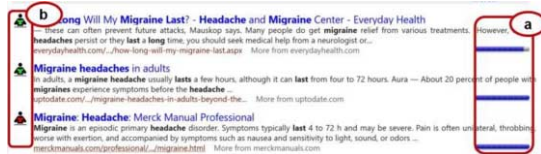


Figure 4 – Trustworthiness and readability filtering for “migraine” with K4E

Quality indicators in a Web browser extension

Using HONcode certified websites as its base, HON developed a Web browser plugin that enriches general search engines (Yahoo, Bing and Google) results with the HONcode seal when health websites are certified. This limits the search only to certified websites and thus reduces its use.

In order to complement the functionalities offered, HON has developed the Health Trust Indicator Web browser extension that includes quality indicators such as the readability level and the results of the automated detection of the HONcode principles on health websites (Figure 5). The readability score is for a given health webpage while the automated HONcode detection indicator is for the whole health website.

This Web browser plugin allows anyone using these general search engines to know for each result the level of trust and of the site providing the information (Figure 5). Then, the user can select trustworthy health pages and avoid information from websites with a low trust score.



Figure 5 – Google results using the Health Trust indicator Web browser extensions

Discussion

The HONcode and Health Trust Indicator Web extensions currently only have a limited impact as people need to be aware of their availability and the quality issue in order to choose to install such services. A solution would be to have certified websites highlighted directly in major search engines without using a Web extension. This tricky issue was explored from 2006 to 2009 using Google Co-Op Topics. HON tagged and labeled HONcode certified webpages that had been manually categorized according to health subject. These webpages were searchable by end users via the beta version of Google Co-Op. However, the results of this service were inconclusive because users did not use it and it was not financially rewarding. It was eventually dropped.

The K4E vertical search engine specializes in health websites, applying specific domain knowledge in the collection of content and in indexing and query formulation. However, only a limited number of people use health search engines as most people favor the convenience of general search engines.

In addition, K4E does not include a HONcode page rank, which could highlight all links from certified websites and offer a popularity score within the limited circle of certified websites. This idea will be studied for implementation within KConnect's further development.

HON also attempted to mobilize the crowd (crowdfunding.org/project/healthwebsiteannotationtest/) to assess websites according to the eight HONcode principles with no success, showing that such a task is too large and complex. However, specific and binary tasks can be achievable, such as to tick yes or no if the date was correctly retrieved by the automated detection system. Thus the next step will be to propose a prototype with a crowdsourcing function associated with the automated HONcode principle detection system. An assessment and evaluation of the task to be performed by the crowd will be conducted as the task should be simplified as much as possible.

HON contributes to webmaster education but seems to have less direct impact on final users. One part of the solution for health and for information in general found on the Internet has to be education and the development of awareness and critical thinking. Information literacy should be taught rapidly and continuously as soon as a child is able to navigate on the Internet [23].

Conclusions

The HONcode is the most used model for the identification of health sites that are transparent and respect quality criteria [24]. However, more effort in terms of access to trustworthy health information is necessary. Twenty years after the inception of HON, there is no solution that is able to address at a large scale the problem of trust on the Internet, particularly for health information. For the past 20 years, HON and its partners have however made a valiant attempt to investigate solutions to the issue of quality health information online. Access to information is mainly a matter of available content, and a matter of search engine algorithms with strong biases as shown in [5]. The so-called bubble filter phenomenon accentuates the problem; another problem is the profiling of the user in search engine results. These drawbacks indicate that substantial room for improvements in currently available common search engines. An alternative approach would create vertical search engines dedicated to health.

An approach combining quasi-automated trust level indicator categorization with manual HONcode certification offers the possibility to cover health websites that have not been evaluated manually. Such reviews can complement the human work involved in evaluation and monitoring of certified health websites.

The HONcode quality assessment tool, which works through a Web browser extension, is straightforward and allows the user to identify if a site is reliable and respects the HONcode principles. The additional automated quality indicators available via the KConnect Web browser extension have shown through usability testing that users appreciate and favor quality online health information and tools when they are aware of them.

Acknowledgements

The research and evaluation activities performed and presented in this article have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644753 (KConnect).

References

- [1] European Commission, Europeans Becoming Enthusiastic Users of Online Health Information, in: *Digital Agenda for Europe*, European Commission, 2014.
- [2] S. Fox and M. Duggan, *Health Online 2013*, Pew Research Center's Internet & American Life Project, 2013.
- [3] S. Fox, *Online Health Search 2006*, Pew Research Center: Internet, Science & Tech, 29 October 2006.
- [4] N. Khomami, Woman dies after taking 'diet pills' bought over internet, in: *The Guardian* April 2015.
- [5] R. W. White, and A. Hassan, *Content bias in online health search*. ACM Transactions on the Web, 2014, 8(4), 25.
- [6] K.M. Griffiths, H. Christensen, Website Quality Indicators for Consumers, in: *J Med Internet Res* (2005);7(5):e55.
- [7] Xin Luna Dong et al., Knowledge-based trust: Estimating the trustworthiness of web sources, in: *Proceedings of the VLDB Endowment* 8.9,(2015):938-949.
- [8] T. Grandison, and M. Sloman, *A Survey of Trust in Internet Applications*, Institute of Electrical and Electronics Engineers (IEEE) Communications Surveys, (2000).
- [9] Y. Gil and D. Artz, Towards content trust of web resources, in: *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, (2007) volume 5, issue 4, 227-239.
- [10] C. Boyer, A. Gaudinat, V. Baujard, A. Geissbuhler. HON Foundation: assessing the quality of health web pages all over the world. In *Medinfo 2007 proceedings Building Sustainable Health Systems* (2007) (p. 1017). IOS Press
- [11] Abbasi, A., F. M. Zahedi, and S. Kaza. 2012. Detecting Fake Medical Websites using Recursive Trust Labeling. *ACM Transactions on Information Systems* 2012.
- [12] P. Sondhi, V. V. Vydiswaran, & C. Zhai, Reliability prediction of webpages in the medical domain. (2012). In *European Conference on Information Retrieval* 219-231. Springer Berlin Heidelberg.
- [13] C. Boyer and L. Dolamic, Automated Detection of HONcode Website Conformity Compared to Manual Detection: An Evaluation, in: *J Med Internet Res*, (2015);17(6):e135.
- [14] A. Hanbury, C. Boyer, L. Dolamic and J. Palotti, *Report on automatic document categorization, trustability and readability*, KHRESMOI Deliverable D.1.6, 2013
- [15] W. M. Silberg, G. D. Lundberg, & R. A. Musacchio. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewer—Let the reader and viewer beware. *JAMA* (1997) 277, 1244–1245.
- [16] I. Bizzi, P. Ghezzi, & P. Paudyal, Health information quality of websites on periodontology. *Journal of Clinical Periodontology*, (2017) 44(3), 308-314.
- [17] AZ Nghiem, Y Mahmoud, R. Som. Evaluating the quality of internet information for breast cancer. *Breast*. (2016); 25:34-37.
- [18] A. U. Joury, M. Alshathri, M. Alkhunaizi, N. Jaleesah, and J. M. Pines, Internet Websites For Chest Pain Symptoms Demonstrate Highly Variable Content and Quality, in: *Academic Emergency Medicine*. (2016)
- [19] C. Boyer, L. Dolamic, P. Ruch and G. Falquet, Effect of the Named Entity Recognition and Sliding Window on HONcode automated detection of HONcode criteria for mass online health content, in: *Proceedings of the 9th International conference on Biomedical Engineering Systems and Technologies (HEALTHINFO)*, 2016, vol. 5: 151-158.
- [20] R. Flesch, A new readability yardstick, in: *Journal of applied psychology*, 32(3), (1948), 221.
- [21] M. De Choudhury, M. Ringel Morris and R. W. White, *Seeking and Sharing Health Information Online: Comparing Search Engines and Social Media*, 1365–76, ACM Press, (2014).
- [22] C. Boyer, J. Hajic, A. Hanbury, M. Kritz, N. Pletneva, P. Schneller, V. Stefanov, Z. Ureskova. 2014 *Report on the extensive tests with final search system KHRESMOI 2014*. Technical Public deliverable, August 2014.
- [23] Eisenberg, Michael B., Carrie A. Lowe, and Kathleen L. Spitzer. *Information literacy: Essential skills for the information age*. Greenwood Publishing Group, 2004.
- [24] P. Wilson, How to find the good and avoid the bad or ugly: a short guide to tools for rating quality of health information on the internet, in: *BMJ: British Medical Journal* 324, no. 7337 (9 March 2002): 598–602.

Arden Syntax Clinical Foundation Framework for Event Monitoring in Intensive Care Units: Report on a Pilot Study

Jeroen S. de Bruin^{a,b}, Julia Zeckl^{a,c}, Katharina Adlassnig^a, Alexander Blacky^d,
Walter Koller^e, Andrea Rappelsberger^b, Klaus-Peter Adlassnig^{a,b}

^a Medexter Healthcare GmbH, Borschkegasse 7/5, A-1090 Vienna, Austria

^b Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

^c University of Applied Sciences Technikum Wien, Höchstädtplatz 6, A-1200 Vienna

^d VAMED-KMB Hospital Management and Operation GmbH, Spitalgasse 23, A-1090 Vienna, Austria

^e University Clinic for Hospital Hygiene and Infection Control, Medical University of Vienna and Vienna General Hospital, Währinger Gürtel 18–20, A-1090 Vienna, Austria

Abstract

The creation of clinical decision support systems has received a strong impulse over the last years, but their integration into a clinical routine has lagged behind, partly due to a lack of interoperability and trust by physicians. We report on the implementation of a clinical foundation framework in Arden Syntax, comprising knowledge units for (a) preprocessing raw clinical data, (b) the determination of single clinical concepts, and (c) more complex medical knowledge, which can be modeled through the composition and configuration of knowledge units in this framework. Thus, it can be tailored to clinical institutions or patients' caregivers. In the present version, we integrated knowledge units for several infection-related clinical concepts into the framework and developed a clinical event monitoring system over the framework that employs three different scenarios for monitoring clinical signs of bloodstream infection. The clinical event monitoring system was tested using data from intensive care units at Vienna General Hospital, Austria.

Keywords:

Expert Systems, Knowledge Bases, Infection Control.

Introduction

Recognition of the benefits and potentialities of information and communication technology in healthcare (eHealth) [1–3] led to political support for healthcare digitization; healthcare institutions were given financial incentives to adopt and make “meaningful use” of electronic health records (EHRs) [4]. Although the term “meaningful use” is quite expansive, it does include the development and use of clinical decision support systems (CDSSs). CDSSs are eHealth systems designed to assist health professionals in clinical decision-making tasks at the point of care.

Clinical event monitors are CDSSs specialized in the delivery of information. A clinical event monitor delivers information to healthcare providers where and when they need it [5]. Generally, a clinical event monitor performs one or more of the following tasks [5]: (a) it issues warnings about adverse events such as potentially harmful drug–drug interactions or complications of treatment, (b) it interprets medical findings, such as

laboratory test results, (c) it provides reminders for immediate or future diagnostic or therapeutic steps, (d) it proposes (alternative) diagnoses or treatment options, and (e) it coordinates complex clinical protocols or workflows.

A substantial number of clinical event monitoring systems have effectively addressed one or more of the aforementioned tasks for a variety of healthcare settings. In the field of infection control, there have been many studies on (semi-) automated systems for the detection and monitoring of healthcare-associated infections [6–8]. Similarly, computerized adverse drug event detection and computerized physician order entry have also been widely researched [9, 10]. The performance of the large majority of systems has been good or excellent. The systems, when measured, proved to be an improvement over traditional or manual methods.

Despite the success of these systems, their use and integration have been limited to their local setting. This is a multifactorial problem. In the present report, we focus on technical and psychological aspects. From a technical point of view, most systems were developed for a specific hospital information system and specific EHRs. Furthermore, they might not always be implemented with established communication standards. As a result, the systems lack interoperability. The effort of porting or recreating the systems outweighs their potential benefits. From a psychological point of view, many systems have only been verified internally. In other words, they have been tested with data from a single healthcare institution. Lacking external validation, the general applicability of the results remains unproven. Moreover, even if a system is verified externally its acceptance by third parties is not guaranteed because the adoption of the system might be perceived as a loss of autonomy [11]. This is especially true of illnesses and adverse events that are not yet fully understood, or which lack consensus regarding their definition or method of detection.

From the above discussion, it follows that the acceptance and dissemination of the system could be improved by providing an interoperable, configurable system. Such a system would use established standards of communication and knowledge representation, thus enhancing its interoperability. A widely known standard for computerized knowledge representation and processing is Arden Syntax [12]. The latter is a programming language for the collection, description, and exchange of medical knowledge in a machine-executable format. Indeed, many of

the tasks performed by clinical monitoring systems have already been modeled in Arden Syntax [13-16]. Improving its acceptance among clinicians would require a knowledge base that could be configured to fit the user's clinical knowledge and experience.

In our view, clinical event monitors are systems that can be composed of standardized configurable building blocks. As such, a limited set of standardized medical knowledge units, which we refer to as the clinical foundation framework, should be available. Based on these, event systems may be constructed and configured according to the wishes of clinical institutions or patient caregivers. These basic blocks of knowledge would be used for preprocessing raw clinical data and determining less complex, clearly defined clinical concepts that are directly measured from objective data and laboratory results. Based on this clinical foundation framework, more complex medical knowledge can then be modeled through the composition and configuration of these basic knowledge blocks.

In the present study, we report preliminary results following the implementation of a clinical foundation framework. We created a clinical event monitoring system that monitors several infection-related clinical concepts based on definitions from internationally respected institutions, such as Centers for Disease Control and Prevention (CDC), Atlanta, USA, and the European Centre for Disease Prevention and Control (ECDC), Stockholm, Sweden. For each of these concepts, we constructed rules in Arden Syntax and integrated them into the clinical foundation framework. Based on a retrospective data analysis with data from intensive care units (ICUs) we show that using knowledge units in the clinical foundation framework as building blocks, we can provide multiple definitions for higher-level clinical concepts.

Methods

Clinical background

We discuss six infection-related clinical concepts included in the clinical foundation framework. These clinical concepts are well-known signs of infection and are used in existing surveillance definitions for infections from the CDC and ECDC. These concepts are fever, leukopenia, leukocytosis, elevated C-reactive protein (CRP), shock, and drop in blood pressure. Definitions of fever, leukopenia and leukocytosis were taken from the "ECDC European surveillance of healthcare-associated infections for intensive care units" protocol, version 1.02 [17]. The definition of elevated CRP was obtained from the CDC National Healthcare Safety Network surveillance definitions for specific types of infection [18]. Definitions of the remaining concepts were constructed by clinical experts. The definitions are listed in Table 1.

Study design, setting, and participants

We conducted a retrospective single-center cohort study on prospectively collected and validated data. The investigation was performed at the Vienna General Hospital (VGH), Austria, which is a 1,933-bed tertiary-care and teaching hospital. Data were collected from patients admitted to one or more of the ICU's at the hospital between 1 January and 31 March 2013. All adult patients (age ≥ 18 years) admitted for at least 24 hours were eligible for the study.

Table 1 – Definitions of clinical concepts modeled in the clinical foundation framework.

Clinical concept	Definition
<i>Fever</i>	Body temperature $> 38^{\circ}\text{C}$
<i>Leukopenia</i>	$< 4,000$ WBC/mm ³ blood
<i>Leukocytosis</i>	$\geq 12,000$ WBC/mm ³ blood
<i>Elevated CRP</i>	CRP > 10 mg/dl blood
<i>Shock</i>	Systolic blood pressure < 1 Heart rate
<i>Drop in BP</i>	BP value in the 37.5% percentile of all averages between systolic and diastolic BP over the last 3 days

Note: WBC, white blood cell; CRP, C-reactive protein; BP, blood pressure.

Data management and sample size

Demographic patient data, as well as clinical and laboratory values, were obtained through systematic interrogation of the Philips IntelliSpace Critical Care and Anesthesia (ICCA) information system, which is in operation at ICUs in the VGH. Interrogation of the data sources using the selection criteria mentioned earlier yielded a total of 984 patient stays.

Knowledge base and data processing

For this project, we reimplemented a part of the knowledge base of Moni (Monitoring of Nosocomial Infections), a fully automated knowledge-based surveillance tool for the identification, monitoring, and reporting of nosocomial (hospital-acquired) infections in ICUs [19].

We used Arden Syntax to implement rules for the clinical infection-related concepts listed in Table 1, as well as rules for data preprocessing and feature extraction. Arden Syntax is a programming language used for representing, processing, and sharing medical knowledge, employed in an executable format by CDSSs to generate alerts, reminders, interpretations, as well as manage messages to clinicians [20]. In an Arden Syntax knowledge base, medical knowledge is divided into medical logic modules (MLMs) [13]; each MLM contains instructions and logic to support at least a single medical decision.

In the clinical foundation framework, MLMs perform one of three types of processing tasks:

- Raw data processing, which deals with importing and processing raw data directly from the structured data source, here the Philips ICCA system.
- Data-to-symbol conversion, which deals with data preprocessing (such as handling missing or contradictory values), and feature extraction (such as calculating mean values or intermediate scores).
- Symbol calculation, which deals with the calculation of basic clinical concepts (e.g., medical symptoms and signs).

In all 17 MLMs were created; seven for raw data import and processing, four for preprocessing and feature extraction, and six for symbol calculation. Table 2 lists these MLMs with a brief description of their task(s).

We used the ARDENSUITE integrated development and test environment (IDE) for the implementation, management, and testing of MLMs in the clinical foundation framework. For the execution of MLMs, we used the ARDENSUITE server [21], to be executed through service-oriented access for client applications.

Table 2 – Medical logic modules part of the clinical foundation framework

MLM name	Task description
Raw data processing	
<i>Temp</i>	Imports body temperature in centigrade over the last 24 hours
<i>ThermoReg</i>	Imports explicit indications of thermoregulation in the last 24 hours, which is performed to cool the patient
<i>Leuko</i>	Imports leukocyte concentrations in G/l
<i>CRP</i>	Imports CRP values in mg/dl
<i>SystBP</i>	Imports systolic blood pressure measurements over the last 24 hours
<i>DiastBP</i>	Imports diastolic blood pressure measurements over the last 24 hours
<i>HeartRate</i>	Imports heart rate measurements over the last 24 hours
Data-to-symbol conversion	
<i>TempMax</i>	Determines the daily maximum body temperature in centigrade
<i>LeukoMax</i>	Determines the daily maximum leukocyte concentration in G/l
<i>CRPMax</i>	Determines the daily maximum CRP in mg/dl
<i>BPPProfile</i>	Determines the blood pressure profile with data over the last 6 hours
Symbol calculation	
<i>TempElev</i>	Determines the presence of fever based on a patient's body temperature
<i>Leukopenia</i>	Determines the presence of leukopenia based on a patient's leukocyte count
<i>Leukocytosis</i>	Determines the presence of leukocytosis based on a patient's leukocyte count
<i>CRPElev</i>	Determines the presence of elevated CRP based on a patient's CRP value
<i>Shock</i>	Determines the presence of (septic) shock based on a patient's systolic BP and heart rate
<i>DropInBP</i>	Determines the presence of a drop in BP based on BP profiles over the last 72 hours

Note: MLM, medical logic module; CRP, C-reactive protein; BP, blood pressure.

Presentation of results

We show how different versions of a system for the detection of clinical signs of infection can be constructed and configured, using basic building blocks from the clinical foundation framework. Based on the data collected from the ICUs at VGH, we show how different setups yield different results.

Results

Of the 984 patients included in this study, 417 were female (42.4%). The youngest was 18 years old, the oldest 92 years; the median age was 61 years, with an interquartile range (IQR) of 24 years. In all 7,573 patient days were recorded during the study period. The length of the hospital stay ranged between two and 93 days, median 4 days, and an IQR of 6 days.

We developed three scenarios, which we model with the clinical foundation framework:

1. *ClinSignsV1*: A straightforward definition of the concept “clinical signs of bloodstream infection” as

specified by the ECDC in [17], involving only the clinical concepts *elevated body temperature*, *leukopenia*, and *leukocytosis*.

2. *ClinSignsV2*: A more complex definition that employs a more comprehensive modeling of the clinical concept of *fever*. In this case, the presence of fever is not only derived from the patient's body temperature, but also from clinical interventions that indirectly indicate the presence of fever, such as the use of cooling packs or blankets (cf., *ThermoReg* in Table 2).
3. *ClinSignsV3*: A definition that extends the *ClinSignsV2* definition by including known markers of infection such as *elevated CRP* and *hypotension*. In this scenario, hypotension is modeled with the clinical concepts of *shock* and *drop in blood pressure*.

Table 3 shows the logical definitions of the clinical concepts used in each scenario. Table 4 shows the number of registered events for relevant clinical concepts and infection symptoms in the clinical foundation framework, and for the clinical concepts listed in Table 3.

Table 3 – Medical logic modules created for various definitions of “clinical signs of bloodstream infection” and their respective logical rules.

MLM name	Definition
Scenario 1	
<i>ClinSignsV1</i>	$TempElev \vee Leukopenia \vee Leukocytosis$
Scenario 2	
<i>Fever</i>	$TempElev \vee ThermoReg$
<i>ClinSignsV2</i>	$Fever \vee Leukopenia \vee Leukocytosis$
Scenario 3	
<i>Hypotension</i>	$Shock \vee DropInBP$
<i>ClinSignsV3</i>	$Fever \vee Leukopenia \vee Leukocytosis \vee Hypotension \vee CRPElev$

Note: MLM, medical logic module; BP, blood pressure; CRP, C-reactive protein.

Table 4 – Symbolic calculation and the resulting number of symptom and scenario events.

MLM name	#Events
Clinical foundation framework	
<i>TempElev</i>	1,394
<i>ThermoReg</i>	4,527
<i>Leukopenia</i>	270
<i>Leukocytosis</i>	2,214
<i>CRPElev</i>	3,606
<i>Shock</i>	2,968
<i>DropInBP</i>	3,217
Scenario 1	
<i>ClinSignsV1</i>	3,268
Scenario 2	
<i>Fever</i>	5,154
<i>ClinSignsV2</i>	5,760
Scenario 3	
<i>Hypotension</i>	4,656
<i>ClinSignsV3</i>	6,835

Note: MLM, medical logic module; CRP, C-reactive protein; BP, blood pressure.

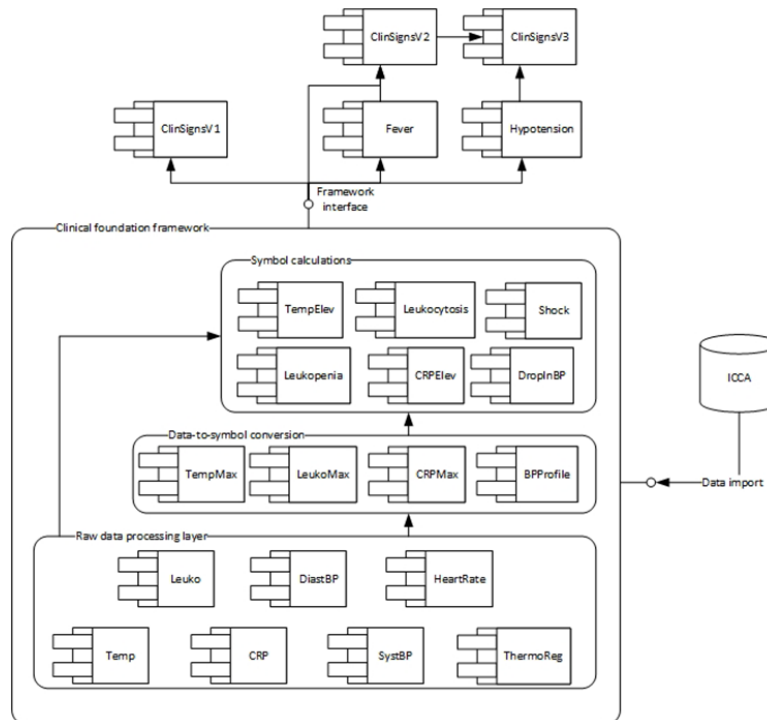


Figure 1 – Graphical depiction of the clinical event monitoring system's architecture. The picture shows the medical logic modules (MLMs) integrated into the clinical foundation framework, and custom-built MLMs for the detection of additional symptoms and infection signs constructed above it. The arrows indicate data dependencies between MLMs or MLM collections. Note: CRP, C-reactive protein; BP, blood pressure.

A graphical depiction of the knowledge base for these clinical concepts (including the clinical foundation framework) is shown in Figure 1.

Discussion

We presented the implementation of a clinical foundation framework in Arden Syntax. The calculation of standardized lower-level clinical concepts directly related to raw clinical data is pre-implemented in a framework of this nature. Consequently, more complex and semantically richer concepts can be calculated by combining elements from the framework with custom implementations. This permits easier and more rapid construction of CDSSs.

The scenarios presented in the Results section all yielded different results. Using the clinical foundation framework, we were able to create different versions of the same clinical concept. This may be useful when the system needs to be implemented for different purposes or different stages of the problem. For example, ClinSignsV1 would be more suited for prospective clinical alerting due to its relatively low number of occurrences, while the more complex ClinSignsV2 would be more suited for retrospective detection of healthcare-associated infections.

The limitations of the study are worthy of mention. First, as the clinical foundation framework is still in its pilot phase, not many MLMs have been implemented so far. Second, we still need to reimplement the systems integrated at VGH in order to

test the framework in a clinical routine. Finally, new systems need to be created and composed in order to assess the ease of construction and improve the performance of the framework and its interfaces.

Several CDSSs have been implemented with Arden Syntax and integrated into clinical routine at VGH, in a variety of clinical specialties, such as nephrology, oncology, and infection control [22]. Inspection of these systems revealed that most of the MLMs in these CDSSs have processing duties performed by the clinical foundation framework, such as raw data processing, data-to-symbol conversion, or symbol calculation. As such, the implementation and configuration of these and similar systems could be simplified by the clinical foundation framework. Furthermore, as the clinical foundation framework grows, an extension of these systems and more complex modeling of symptoms, signs, interpretations of laboratory test results, clinical findings, diseases, therapies, adverse events, quality measures, etc. will become easier.

Conclusion

We created a clinical foundation framework, based on which clinical event monitoring systems can be constructed through combination and configuration. Using the framework, CDSSs can be created more rapidly and configured according to the specific needs of healthcare institutions and patients' caregivers.

References

- [1] A. Dobrev, T. Jones, V. Stroetmann, K. Stroetmann, Y. Vatter, and K. Peng, *Interoperable eHealth is Worth it—Securing Benefits from Electronic Health Records and ePrescribing*, 2010. Available at: http://www.ehr-impact.eu/downloads/documents/ehr_impact_study_final.pdf, last access: 25 October 2016.
- [2] R.L. Duffy, S.S. Yiu, E. Molokhia, R. Walker, and R.A. Perkins, Effects of Electronic Prescribing on the Clinical Practice of a Family Medicine Residency, *Fam Med* **42** (2010), 358–363.
- [3] L.M. Kern, Y. Barron, R.V. Dhopeswarkar, A. Edwards, and R. Kaushal, Electronic Health Records and Ambulatory Quality of Care, *J Gen Intern Med* **28** (2013), 496–503.
- [4] E.V. Murphy, Clinical Decision Support: Effectiveness in Improving Quality Processes and Clinical Outcomes and Factors that may Influence Success, *Yale J Biol Med* **87** (2014), 187–197.
- [5] G. Hripcsak, P.D. Clayton, R.A. Jenders, J.J. Cimino, and S.B. Johnson, Design of a Clinical Event Monitor, *Comput Biomed Res* **29** (1996), 194–221.
- [6] J.S. de Bruin, W. Seeling, and C. Schuh, Data Use and Effectiveness in Electronic Surveillance of Healthcare Associated Infections in the 21st Century: A Systematic Review, *J Am Med Inform Assoc* **21** (2014), 942–951.
- [7] R. Freeman, L.S. Moore, L. Garcia Alvarez, A. Charlett, and A. Holmes, Advances in Electronic Surveillance for Healthcare-Associated Infections in the 21st Century: A Systematic Review, *J Hosp Infect* **84** (2013), 106–119.
- [8] J. Leal and K.B. Laupland, Validity of Electronic Surveillance Systems: A Systematic Review, *J Hosp Infect* **69** (2008), 220–229.
- [9] S.M. Handler, R.L. Altman, S. Perera, J.T. Hanlon, S.A. Studenski, J.E. Bost, M.I. Saul, and D.B. Fridsma, A Systematic Review of the Performance Characteristics of Clinical Event Monitor Signals used to Detect Adverse Drug Events in the Hospital Setting, *J Am Med Inform Assoc* **14** (2007), 451–458.
- [10] K.M. Marasinghe, Computerised Clinical Decision Support Systems to Improve Medication Safety in Long-Term Care Homes: A Systematic Review, *BMJ Open* **5** (2015), e006539.
- [11] Z. Walter and M.S. Lopez, Physician Acceptance of Information Technologies: Role of Perceived Threat to Professional Autonomy, *Decis Support Sys* **46** (2008), 206–215.
- [12] Health Level Seven, *Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10)*, 2015. Available at: http://www.hl7.org/implementation/standards/product_brief.cfm?product_id=372, last access: 25 October 2016.
- [13] G. Hripcsak, Writing Arden Syntax Medical Logic Modules, *Comput Biol Med* **24** (1994), 331–363.
- [14] C.Y. Jung, J.Y. Choi, S.J. Jeong, K. Cho, Y.D. Koo, J.H. Bae, and S. Kim, Transformation of Arden Syntax's Medical Logic Modules into ArdenML for a Business Rules Management System, *Artif Intell Med* (2017), doi: 10.1016/j.artmed.2016.03.005.
- [15] S. Kraus, C. Drescher, M. Sedlmayr, I. Castellanos, H.U. Prokosch, and D. Toddenroth, Using Arden Syntax for the Creation of a Multi-Patient Surveillance Dashboard, *Artif Intell Med* (2017), doi: 10.1016/j.art-med.2015.09.009.
- [16] A. Seiting, A. Rappelsberger, H. Leitch, M. Binder, and K.-P. Adlassnig, Executable Medical Guidelines with Arden Syntax—Applications in Dermatology and Obstetrics, *Artif Intell Med* (2017), doi: 10.1016/j.art-med.2016.08.003.
- [17] European Centre for Disease Prevention and Control (ECDC), *European Surveillance of Healthcare-Associated Infections in Intensive Care Units, HAI-Net ICU Protocol, Protocol Version v1.02*, 2015. Available at: <http://ecdc.europa.eu/en/publications/Publications/healthcare-associated-infections-HAI-ICU-protocol.pdf>, last access: 16 November 2016.
- [18] Centers for Disease Control and Prevention (CDC), *CDC/NHSN Surveillance Definitions for Specific Types of Infections* 2016. Available at: http://www.cdc.gov/nhsn/pdfs/pscmanual/17pscnosinfdef_current.pdf, last access: 16 November 2016.
- [19] J.S. de Bruin, K.-P. Adlassnig, A. Blacky, H. Mandl, K. Fehre, and W. Koller, Effectiveness of an Automated Surveillance System for Intensive Care Unit-Acquired Infections, *J Am Med Inform Assoc* **20** (2013), 369–372.
- [20] M. Samwald, K. Fehre, J. de Bruin, and K.-P. Adlassnig, The Arden Syntax Standard for Clinical Decision Support: Experiences and Directions, *J Biomed Inform* **45** (2012), 711–718.
- [21] Medexter Healthcare, *ArdenSuite – Medical Knowledge Representation and Rule-Based Inference Software with Arden Syntax*, 2015. Available at: <http://www.medexter.com/arden-syntax/arden-syntax>, last access: 8 November 2016.
- [22] C. Schuh, J.S. de Bruin, and W. Seeling, Clinical Decision Support Systems at the Vienna General Hospital Using Arden Syntax: Design, Implementation, and Integration, *Artif Intell Med* (2017), doi: 10.1016/j.artmed.2015.11.002.

Address for correspondence

Univ.-Prof. DI Dr. Klaus-Peter Adlassnig, FACMI Section for Artificial Intelligence and Decision Support
 Center for Medical Statistics, Informatics, and Intelligent Systems
 Medical University of Vienna
 Spitalgasse 23, 1090 Vienna, Austria
 E-mail: klaus-peter.adlassnig@meduniwien.ac.at and
 Medexter Healthcare GmbH Borschkegasse 7/5, 1090 Vienna, Austria
 E-mail: kpa@medexter.com

Design of an Intelligent Nursing Clinical Pathway and Nursing Order Support System for Traditional Chinese Medicine

Bao-Fen Ding^a, Polun Chang^b, Ping Wang^c, Hai-Ting Li^d, Ming-Chuan Kuo^{b,e}

^a Nanjing Chest Hospital, *Medical College of Southeast University, Nanjing, Taiwan, China*

^b Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

^c He Ren (Tianjin) Science and Technology Co., Ltd, Tianjin, China

^d Henan Province Luoyang Orthopedic Traumatological Hospital, Luoyang, China

^e Cathay General Hospital, Taipei, Taiwan, China

Abstract

With an in-depth analysis of nursing work in 14 hospitals over a period of two years, one unique total nursing information system framework was established where the nursing clinical pathways are used as the main frame and the nursing orders as the nodes on the frame. We used the nursing order concept with the principles of nursing process. A closed-loop management model composed of the nursing orders was set up to solve nursing problems. Based on the principles of traditional Chinese medicine, we further designed an intelligent support module to automatically deduct clinical nursing pathways to promote standardized management and improve the quality of nursing care. The system has successfully been implemented in some facilities since 2015.

Keywords:

Medicine, Chinese Traditional; Nursing Process; Disease Management

Introduction

One of the important aspects of modern nursing science is the systematic holistic nursing care. It has not only been affected by the information technology, but also laid the foundation for nursing informatics, which was pointed out by a Nursing informatics pioneer, Dr. Hannah, in 1985 [1]. Medicine and health in China has developed rapidly in the past 10 years except nursing informatics. We conducted a thorough investigation in 14 top hospitals among five Chinese cities and found that there is no complete and effective clinical nursing information system, especially for the Chinese medical hospitals. Therefore, we set the goals of the study as follows:

- To create a standardized nursing clinical framework with Nursing Clinical Pathway (NCP) as the main frame
- To create core elements of the system with nursing order
- To guide the whole nursing process of the closed-loop management paradigm through nursing orders
- To introduce an intelligent decision support system (DSS) module to automatically deduct procedures on the principles of Chinese Medicine [2]

Methods

1. Extensive and In-depth Research Work

Overview of our survey on nursing clinical work follows:

1.1 Time Period: May 2012 - May 2014.

1.2 Targets: Top 14 hospitals in five cities located in different areas: Beijing, Tianjin, Nanjing, Xiamen, Luoyang are selected with more emphasis on Luoyang hospital.

1.3 Team Members: Three medical and health information specialists from mainland China and Taiwan, six directors from Nursing and Information Department, and four Software Engineers.

1.4 Process: On the basis of general survey, comprehensive investigation in Luoyang orthopedic hospital was carried out. First, 'Xiang bi' disease (Cervical Radiculopathy) and other three kinds of diseases were chosen, 200 electronic medical records and 551 nursing recording sheets were read. Second, Eight seminars were held. Third, 20 on-site investigations of wards were carried out. Fourth, 22 teleconference meetings were held. A comprehensive and complete information was collected [3].

2. Deep Analysis of Collected Materials and Data and Proposal of the Design Framework

2.1 Analysis results of 200 electronic medical records

We found that the degree of similarity of nursing work for the same kind of disease was 87%. Also, after cluster analysis of the top 30 nursing tasks of the highest frequency, we found that 85% were about hospitalization in the same period. For example, "Admission Assessment" happened on the first day, "Cervical Traction" happened between days 5 – 25. These were consistent with the "Rules of Diagnosis and Treatment of Cervical Spondylosis" that provided the theoretical basis for the design NCP.

2.2 Analysis results of 50 'Xiang bi' disease (Cervical Radiculopathy) electronic medical recording sheets

Only 14% of nursing tasks were derived by Nursing Orders. For example, sheets like "Pain Assessment" and "PIO Recording" were filled out with traditional habits, the regular rounds of "Intravenous Injection" were often ignored. As a result, we put forward a concept of "Nursing Order". Namely, each kind of nursing behavior should be derived by Nursing Order to standardize the content, process, and results of nursing work.

2.3 Analysis results of 551 nursing recording sheets

There were more than 82 kinds of nursing recording sheets, and in one sheet, such as “Admission Recording Sheet”, as many as 128 items were included, and the duplication of content and data were in a high proportion, which wasted a lot of time and energy of nurses [4, 5]. During the analysis of “Intravenous Administration Recording Sheets”, we found that there was an outstanding feature of nursing work, namely, cycle and repetition. Based on this feature, a closed-loop nursing and management model was designed.

Taking long-term intravenous administration as an example, we divided it into eight nodes, and standardized the administration place, treatment method, administering nurse, and treatment outcome of each node as per its information flow. For example, the first node “Decomposition of Doctor’s Order”; the administration place was nursing station; the treatment method was to automatically transfer each order into the concrete operation content; the executor was the nurse on duty. Another example was the last node “Infusion Result Record”. There were 3 kinds of “evaluations”: If the evaluation was “Effective”, then continue with the cycle for tomorrow. If the evaluation is “Adverse reactions”, then terminate and quit. If the evaluation was “Cure or Stop Administration”, then end such long-term administration. See figure-1:

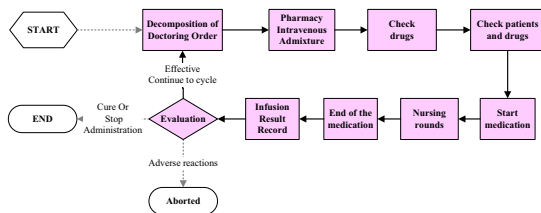
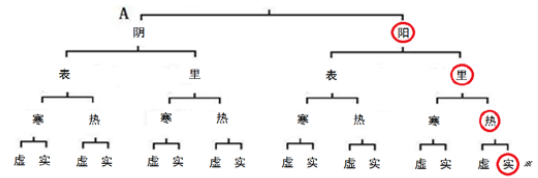


Figure 1 – Closed Loop Nursing Management Model of Intravenous Administration

3. Exploring Decision Support System on the Basis of Traditional Chinese Medicine Theory

The principle of diagnosis and treatment of dialectical therapy in traditional Chinese Medicine is completely different from Western Medicine. The core principle of traditional Chinese Medicine is “Yin and Yang” theory, which is the dialectical treatment mainly based on “Yin and Yang” and the 4 groups of opposite syndrome. They are Yang \longleftrightarrow Yin, Biao \longleftrightarrow Li, Shi \longleftrightarrow Xu, and Re \longleftrightarrow Han [6]. Doctors and nurses would analyze clinical data of patients through four types of dialectical analysis and qualitative analysis of each group to obtain the diagnosis, as shown in Figure 2 [1]. If the analysis result is in the red circle position, the diagnosis is “Yang-Li-Re-Shi Symptom”.

The dialectical analysis method and the binary language of computing follow a strikingly similar path. If we use “1” to stand for Yang, Biao, Shi and Re, and use “0” to stand for Yin, Li, Xu and Han, we can get the map of Figure 3 [1]. “1011” can be used to express the last diagnosis in red circle – “Yang-Li-Re-Shi Symptom”. Above idea is one of the methods and principles we adopted to develop traditional Chinese Medicine DSS.

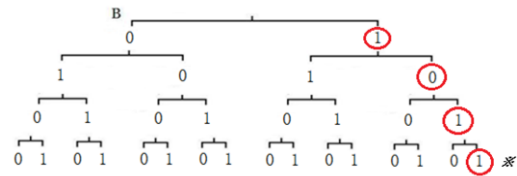


Remark:

Chinese word	English Code	Comment
阴	Yin	positive
阳	Yang	negative
表	Biao	exterior
里	Li	interior
寒	Han	cold
热	Re	heat
虚	Xu	deficiency
实	Shi	excess

※ Yang- Li- Re - Shi Symptom

Figure 2 – Schematic Diagram of Treatment “Principle of Traditional Chinese Medicine Dialectical Therapy”



※ 1011

Figure 3 – Schematic Diagram of Computer Binary Expression

4. Data Modeling Methods

4.1 Nursing Information Data Abstracted in Structured Form

General structure storage was simple and accurate, but separate settings for each type of information needed to be set up. In the nursing information system, intensive maintenance effort was caused by the diversity of information. Besides, diversity of data would cause poor generality of system. Therefore, we abstracted the collected data of nursing index, and stored them into a standard model to ensure data standardization and universality, which could also satisfy the diversity of data and the demand of reuse. This storage mode could serve subsequent DSS, remainder of plans, and result analysis better. It could also ensure users to gradually formulate their own knowledge base of their own business and processes for later use [7].

4.2 Decision Support System Algorithm Based on Bayesian Theorem

On account of the process of closed loop nursing management, we used the main attribute features deduced from the various nursing assessment content of patients. Based on structured data storage, we chose Naive Bayesian classification method in the Bayes classifier model as the foundation of intelligent algorithm [8, 1]. Through the analysis of results of evaluation, subsequent nursing plan was proved to be feasible. Combining recommended nursing plans and subjective judgment; the high accuracy rate of nursing plan was available to users and could help users execute better patient care.

Results

1. The Creation of Traditional Chinese Medicine NCP

The concept of NCP: Creating nursing diagnosis classification according to the disease type or patient condition first, and then formulating regulatory nursing plan for each disease type entailing standardized nursing process [9].

The purpose of using NCP was to establish the main framework of nursing work, and applying all huge and tedious nursing tasks in standard and orderly manner. Standardization meant that it was based on the knowledge base of nursing and evidence-based medicine, providing a common solution to the given nursing problem. Ordering meant that it arranged nursing work automatically according to time series of patients. NCP from different time periods referred to the nursing work derived from traditional Chinese medicine in accordance with its diagnosis principles, special laws, and unique methods. Figure 4 is the NCP of 'Xiang bi' disease (Cervical Radiculopathy) in traditional Chinese medicine.



Figure 4 – NCP of 'Xiang bi' Disease (Cervical Radiculopathy) in Traditional Chinese Medicine

As shown in Figure 4, the first row is the length of stay of the patient, which forms the main timeline of the pathway. The second row shows the four nursing stages. The third row is the nursing order (like evaluation sheets) derived automatically. When we click the current nursing order on the right side, evaluation sheets for each item of nursing content will display automatically.

2. To Create a "Nursing Order"

The clinical nursing work of patients included a large number of nursing content such as, the patient admission assessment, and regular turning over of patients. These tasks were not clearly defined in the past – they might have been implemented as a habit silently or hidden in the corner of a form or even ignored or forgotten.

The concept of Nursing Order: The nurse gave nursing plans or orders according to the patient's nursing problems. Every nursing order referred to a specified nursing behavior with unambiguous content, creation time, and execution time.



Figure 5 – Nursing Order of Traditional Chinese Medicine "admission assessment"

Nursing Order specified the corresponding recording sheet. For example, Figure 5 shows admission assessment.

Nursing Order fully defines nursing tasks clearly and formally for the first time. It promotes nurses' compliance to standard work practice, ensures the quality of care, and decreases nursing errors.

3. To Design Closed-loop Nursing Quality Management Model

Now nursing work reflected its internal rules and processes. Nursing work had one notable feature that was used repeatedly until the disease was resolved or cured [5]. We introduced the closed-loop management of long-term intravenous administration in Methods 2.3. It contains eight nursing operation links that formed a closed-loop management. Nurse could process in turn according to the reminder on handheld PDA, and input data accordingly. Figure 6 is the operation interface of handheld PDA.



Figure 6 – The Operation Interface of Handheld PDA

4. To Create NCP Decision Support System

The basic framework of systematic approach to holistic nursing care was nursing process. It consisted of five aspects, which were arranged in sequence and interacted with each other and formed a closed loop: Nursing Assessment → Nursing Diagnosis → Nursing Planning → Nursing Measures → Nursing Evaluation. Therefore, the nursing process was information collection, transmission, analysis, processing knowledge application and accumulation process [10, 11]. The nursing pathway regulated all the nursing workflow and information processing procedure, and the closed-loop management standardized the process of each nursing task, thus providing a basis for DSS.

Based on the above ideas, we designed DSS for NCP. Taking pressure ulcer care as an example [12], all the information

flow was divided into five nodes, and we regulated the execution place, treatment method, executing nurse, and result judgement of each node. As shown in Figure 7:

- "Assessment": Nurse collected information of sacral ulceration through interrogation and physical examination and filled in "Pressure ulcer assessment form".
- "Diagnosis": According to the assessment the system concluded that "sacral pressure sores" diagnosis.
- "Plan": According to the diagnosis, the system automatically provided optional nursing measures, such as debridement, dressing change, regular turn over and generating "pressure sores treatment record".
- "Implementation": Nurse completed the nursing care and filled out the treatment list.
- "Evaluation": Nurse filled in the treatment result after completion of each nursing task. System would automatically determine the three kinds of treatments according to the efficacy. Effective – automatically generated the nursing order of next day and enter into the minor-cycle of "Implementation" → "Evaluation". Invalid – automatically produced the nursing order of "Pressure ulcer assessment" on the next day, and re-entered the big loop. Cured – exit the loop and stop "pressure ulcer care" on the next day.

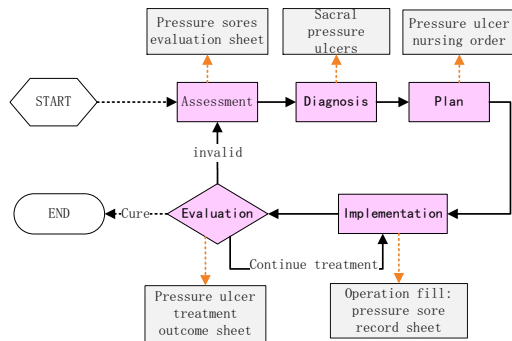


Figure 7 – Pressure Ulcer Care DSS Flow Diagram

5. To Establish NCP Decision Support System Based on Traditional Chinese Medicine Theory

Traditional Chinese medicine has a special theory and methods with traditional diagnosis and treatment. It is formed and developed in China over thousands of years, and is greatly different from Western Medicine. Again, let's take the pressure ulcer as an example. According to different symptoms and signs, the traditional Chinese medicine divided pressure ulcer into four different types and every type is divided into four different periods. Different types and periods have different traditional Chinese medicine nursing methods, for example, Chinese medicine fumigation, moxibustion and so on [13].

The difficulty in establishing the NCP decision support system based on traditional Chinese medicine theory is the cognition, understanding and informational expression of various theories of traditional Chinese medicine. Chinese medicine DSS consists of four parts:

- Knowledge database was built according to the Chinese medicine theory, such as table of clinical symptom.
- Establish the mathematical model of knowledge reasoning of traditional Chinese medicine, simulate the inference process of the syndrome differentiation and treatment of tradi-

tional Chinese medicine, and establish the statement of "Yin and yang theory" in the "method" section.

- Design of the program, to achieve the above mathematical model with the computer language.
- Clinical verification to verify its compliance rate with a large number of cases.

6. Practical Application

The IRB committee with all five votes approved this study.

We have applied our work to three hospitals in a year, and have achieved good results (Table 1). Nurses could choose NLP based on patient's diagnosis, and the NLP automatically displayed physician and nurse orders. Nurses could then follow all procedures and complete documentation. The NLP DSS was used to monitor nurse's tasks and it provided reminders for corrections to assure the care quality. The system had passed the international HIMSS Level 6 assessment in two hospitals.

Table 1 - Application Effect of Contrast

Item	Before Use	After Use	Effect
Total Forms (species)	82	18	↓ 64
Fill out forms time (min)	17	4	↓ 13
Manual input data (%)	85	0	↓ 85
Input by PDA (%)	15	95	↑ 80
Nurse satisfaction (%)	7	92	↑ 85

Discussion

1. To Explore the Law of Nursing Work with Thorough Investigation and Study

We had a profound understanding, in order to find out the objective law of the nursing work, and the various factors affecting the law, we had to do long-term, in-depth investigation and study all the data again and again. In addition, in order to realize the management of nursing information, we had to do innovative design, to explore the way of expressing the objective law, and to control influencing factors with information technology [14].

2. To Create a Standardized Nursing Clinical System Framework

We created an overall framework of the nursing work with NCP as the main line and nursing order as the node. The pathway is the nursing work according to a patient's length of stay. The nursing order is the nursing instruction of a certain point in the pathway, and corresponding to a specific nursing form. In this way, we summarized the whole nursing work to the overall framework in a standardized and well-organized way that could guide the nursing work correctly.

3. Meaning of Creating Nurse Order

We only had the doctor order that was the treatment plan or instruction given by a doctor. But "nursing order" could not be searched in document retrieval. Its creation has the following main significance [15]:

Firstly, nursing order refers to instructions of nursing work – each nurse order refers to a specific nursing behavior. It is the first time that there is a nursing order dictionary that shows name and content of nursing work. It is convenient for scheduling, inspection, and statistical analysis.

Secondly, we designed a nursing process that is a closed-loop system with five links with five specific nursing orders with the extension of pathway, cycle, and spiral propagation until nursing problem is solved in the NCP.

Thirdly, nursing order and nursing records standardized nurse work, avoided negligence and omission. This is conducive to improve quality of nursing and to prevent accidents.

4. Importance of Decision Support System for Nursing Pathway

NCP is based on the statistical analysis of medical knowledge database and evidence based medicine. It reflects the correct treatment route with a high probability. Since it is the result of a high probability, it is impossible to adapt to disease changes of every patient on every day. This does not conform to the currently new concepts of "precision medicine treatment", "personalized treatment". This is also poses problem in applying correct clinical pathway.

DSS can deduct the nursing order for the next time based on the evaluation of the current nursing result automatically and gives a choice for nurse for treatment and thus achieves the "precision of medical treatment" and "personalized therapy".

The knowledge database of nursing pathway and closed-loop management—and in particular, the rule database—is relatively simple and clear, so it is easy to establish the mathematical model of knowledge reasoning, and to design DSS.

5. Difficulties in Creating NCP Based on Chinese Medicine Theory

Although we had completed the exploration of the 'Xiang bi' disease (Cervical Radiculopathy) and other four diseases of traditional Chinese medicine in NCP, there are still a lot of problems. Method of information processing is not established for many principles due to broad and profound traditional Chinese medicine. In addition, the standardization of traditional Chinese medicine terminology is still in progress. We and the other Chinese medicine information experts are waiting for this long-term effort. The principles of creating CNP in West or East medicine are all the same but the difference mainly comes from the medical principles and knowledge.

Conclusion

We had a general survey of the nursing work in 14 hospitals, including a 2-year in-depth investigation at Luoyang Orthopedic Hospital. We achieved many objectives during this work and our intended target turned more realistic after 1-year clinical application.

Firstly, we created an overall framework of the nursing information system with NCP as the main line and nursing order as the node. Secondly, we completed the design and development of nursing pathway. Thirdly, we created a concept called "nursing order", edited the nursing orders data dictionary, and used it to standardize the nursing work. Fourthly, through the closed-loop management model formed by nursing order, we have advanced along the NCP' spiral path, and solved various nursing problems. Fifthly, we designed the intelligent NCP function that is conducive to the principle of "precision of health care" and to the improvement of quality of nursing.

We developed a NCP decision support system based on the principle of traditional Chinese medicine and explored the ideas and methods of research and development. But there are still many problems due to the nature of broad and profound traditional Chinese medicine. Its standardization still needs to be improved; we, therefore, have scheduled a special R&D program for the next three years. In addition, we are promoting this achievement to more hospitals and specialists to enhance the perfection and maturity of the product.

References

- [1] Ding Baofeng. Medical Informatics[M]. Nanjing: Southeast University press, 2009.9.
- [2] Kerstin Denecke, Cord Spreckelsen. Personalized Medicine and the Need for Decision Support Systems[J]. Data and Knowledge for Medical Decision Support, 2013, 186:41-45.
- [3] Gardner RM et al. Core content for the subspecialty of clinical informatics. JAMIA, 2009 16(2):153-7.
- [4] Jiang Hui, Ma Lili, Li Wangying, et al. Design and application of nursing documentation based on clinical pathway and high quality nursing[J]. Chinese Journal of nursing, 2011, 46(9):3-6.
- [5] Dai Xiangjun, Ye Zhou, Zhang Lei, et al. Building and Application of Intelligent Knowledge Base Based on the Whole Link of Diagnosis and Treatment[J]. Chinese Digital Medicine, 2014, 9(5):18-20.
- [6] Li Zhaoguo. On the concept, principles and methods of the international standardization of English translation of Chinese medical terms[J]. Chinese Translators Journal, 2008, (4).
- [7] Charlotte Weaver, Ann O'Brien, et al. Transforming Clinical Documentation in EHRs for 2020: Recommendations from University of Minnesota's Big Data Conference Working Group[J]. Nursing Informatics 2016, 2016, 225:18-22.
- [8] Edward H. Shortliffe and James J. Cimino. Biomedical Informatics: Computer Applications in Health Care and Biomedicine. 3rd ed. New York: Springer, 2006.
- [9] Hu Bin, Ma Jun, Xu Yanling, et al. Implementation of Nursing Decision Support System Oriented toward Clinical Pathway Application[J]. Hospital Digitalization, 2013, 34(6):2-4.
- [10] Marko Knežević, Vladimir Ivančević, Ivan Luković. A Context-Sensitive Support System for Medical Diagnosis Discovery based on Symptom Matching[J]. Intelligent Decision Technologies, 2013, 255:1-10.
- [11] Myoung-Rye Bong, Young-Seon Lee, Kwang-Ok Park, et al. A Clinical-decision Support System for the Nursing Process in the ENR[J]. Connecting Health and Humans, 2009, 146:815-816.
- [12] Wang Weixing, Huang Jin, Li Lezhi, et al. Design and Application of Inpatient Pressure Sore Management System Based on Internet[J]. Chinese Digital Medicine, 2016:53-55.
- [13] Li jinsong et al. Biomedical semantic technology[M]. Zhejiang: Zhejiang University Press, 2012.8
- [14] Rodrigo Jensen, Christiane Pereira Martins Casteli, Rika Miyahara Kobayashi, et al. Informatics Competencies in Nursing Management[J]. MEDINFO 2015: eHealth-enabled Health, 2016, 216:1013-1013.
- [15] Jin Xinzheng, Ding Baofeng, et al. Hygienic Information Management Syst[M]. Beijing: People's Medical Publishing House, 2009.1.

Address for correspondence

Bao-Fen Ding
Professor, Nanjing Chest Hospital, Medical College of Southeast University, Nanjing
Tel: +86-02583721882
E-mail: dingbaofen@sina.com

Predicting Length of Stay for Obstetric Patients via Electronic Medical Records

Cheng Gao^a, Abel N. Kho^b, Catherine Ivory^c, Sarah Osmundson^d, Bradley A. Malin^{a, e, f}, You Chen^a

^aDept. of Biomedical Informatics, School of Medicine, Vanderbilt University, Nashville, TN, USA

^bInstitute for Public Health and Medicine, Northwestern University, Chicago, IL, USA

^cSchool of Nursing, Vanderbilt University, Nashville, TN, USA

^dDept. of Obstetrics and Gynecology, School of Medicine, Vanderbilt University, Nashville, TN, USA

^eDept. of Electrical Engineering & Computer Science, School of Engineering, Vanderbilt University, Nashville, TN, USA

^fDept. of Biostatistics, School of Medicine, Vanderbilt University, Nashville, TN, USA

Abstract

Obstetric care refers to the care provided to patients during ante-, intra-, and postpartum periods. Predicting length of stay (LOS) for these patients during their hospitalizations can assist healthcare organizations in allocating hospital resources more effectively and efficiently, ultimately improving maternal care quality and reducing costs to patients. In this paper, we investigate the extent to which LOS can be forecast from a patient's medical history. We introduce a machine learning framework to incorporate a patient's prior conditions (e.g., diagnostic codes) as features in a predictive model for LOS. We evaluate the framework with three years of historical billing data from the electronic medical records of 9188 obstetric patients in a large academic medical center. The results indicate that our framework achieved an average accuracy of 49.3%, which is higher than the baseline accuracy 37.7% (that relies solely on a patient's age). The most predictive features were found to have statistically significant discriminative ability. These features included billing codes for normal delivery (indicative of shorter stay) and antepartum hypertension (indicative of longer stay).

Keywords:

Length of Stay; Electronic Health Records; Obstetrics

Introduction

Electronic medical record (EMR) systems have been widely adopted in the United States (US) and abroad [1-4]. These systems enable a substantial amount of data to be captured during the routine practice of healthcare organizations (HCOs) [2-5]. This information is quite heterogeneous, including structured diagnoses, medication regimens, laboratory test results and vital signs, as well as un- or semi-structured clinical narratives. The data stored in EMRs is increasingly recognized for its ability to support numerous activities, such as clinical decision making [6], patient safety improving [7-8] and discovery-driven biomedical research [2-4].

Currently, some of the most challenging healthcare environments to manage for safety are those associated with maternity. Over the past several decades, the maternal mortality ratio (MMR) has risen dramatically in the US. MMR has doubled from 7.2 deaths of mothers per 100,000 live births in 1987 to 14 in 2015 [9]. At the same time, obstetric care is the most common and costly type of hospital care for all payers in the US [10-12]. Prediction of the length of stay (LOS) for obstetric patients during their hospitalization can

help unit managers and administrators make decisions about hospital resource allocation - enabling obstetric care improvement before, during and after childbirth. This is notable because better organized care can reduce the morbidity and mortality of women, as well as newborn babies [11;12], while reducing maternity-related costs. The incorporation of an accurate estimate of LOS in counseling discussions may mitigate anxieties over the uncertainty of a hospital stay as well as prepare for discharge to home or elsewhere [13]. This is important both for obstetric patients and their families who often inquire about the expected duration of a hospitalization.

Previous research has focused on characterizing the factors that lead to LOS variation in general. LOS has, for instance, been shown to be influenced by a patient's demographics (e.g., age), socioeconomic status (e.g., income, education, and occupation), insurance types (e.g., commercial, private, and Medicaid and Medicare) and severity of illnesses [14-16]. LOS has further been shown to be affiliated with HCO-specific factors, such as physicians' work efficiency [5,17], climate [18] and the availability of professional language interpretation services [19]. However, the complex relationships between these factors further exacerbate the complexity of LOS prediction. Thus, it is challenging to build LOS prediction models that rely solely on expert knowledge and information ascertained at the time of a patient's admission to a hospital.

In recognition of these limitations, this paper presents a pilot study on the feasibility of a patient's historical diagnoses, as documented in an EMR, for LOS predictive models. This study is predicated on the hypothesis that LOS is related to a patient's medical history. To investigate this hypothesis, we study three years worth of historical diagnosis codes (prior to their most recent admission) for patients on an obstetric service at Northwestern Memorial Hospital (NMH) in Chicago, Illinois, USA. Specifically, we extracted EMR data in the form of International Classification of Diseases, ninth revision (ICD-9) [20] codes and designed a machine learning framework to predict LOS. The results indicate that prediction of LOS within 12 hours can be achieved with almost 30% greater accuracy than the baseline model that relies solely on the patient's demographics at the time of admission. In addition, we show that certain billing codes are statistically significant in their predictive capability, which suggests they are ripe for further investigation and transition into clinical decision support.

Methods

Figure 1 provides the EMR data and analytics workflow adopted for this investigation. First, the ICD-9 codes and LOS for patients are extracted from the EMR. These are subsequently applied to train and test a predictive model. Finally, the most discriminant ICD-9 codes are prioritized and assessed for statistical significance.

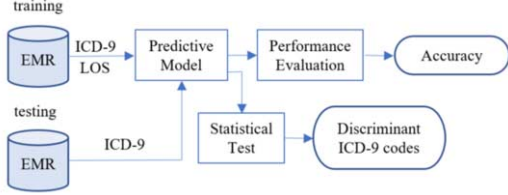


Figure 1 – The process by which the LOS predictive model is composed and discriminative features are discovered.

Dataset

The dataset was drawn from the Cerner inpatient EMR system in place at NMH from July 2007 to July 2011. It includes the following patient-specific features: 1) demographics (e.g., age), 2) encounter information (e.g., admission and discharge date), 3) diagnosis (e.g., billing codes) assigned to an encounter, and 4) clinical (e.g., obstetrics) service to which the patient was assigned. In total, there were 9188 inpatients in the dataset with 1849 distinct ICD-9 codes. We consider all inpatients on the obstetric service during 2010-2011 for prediction and rely on EMR data between 2007-2009 as features for our models.

The LOS for an encounter was calculated as the hourly difference between admission and discharge. We use a patient's age as a baseline prediction for LOS. Table 1 summarizes the average number of ICD-9 codes for the investigated patients in one-, two- and three-years of EMR data, the average age of the investigated patients, and the average LOS for these patients on the obstetric service during the 2010-2011 period.

Table 1 – Summary Statistics for ICD-9 codes, age and LOS in the 2010-2011 period

	# of ICD-9 codes			Age	LOS
	1 year	2 years	3 years		
Mean	4.3	5.5	6.4	31.8	72
Min	1	1	1	14	1.6
Max	60	80	90	68	1311

The distribution of inpatients on LOS is shown in Figure 2. The LOS for the majority (74%) of obstetric patients ranges from 48 to 96 hours.

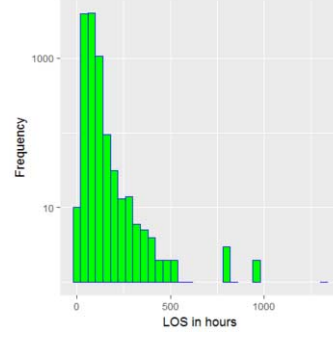


Figure 2 – LOS Frequency distribution for study subjects.

Predictive Model

We adopted a random forest model to predict a patient's LOS according to their historic assigned ICD-9 codes. We rely on a random forest because it is a useful ensemble approach for regression and classification. Specifically, the average LOS from all the trees is used for prediction.

We model the data as a matrix, as shown in Equation (1). Let n be the number of patients and m be the number of unique ICD-9 codes. In this matrix, each row represents a specific patient and each column is a specific characteristic of the patient. The first column is a patient's LOS (continuous variable) and the rest of the columns are the ICD-9 codes for each patient. To mitigate the influence of repeat visits for patients, we treat each ICD-9 code as a binary variable, such that it is set to 1 if the patient received this diagnosis at least once and 0 otherwise.

$$\begin{bmatrix} LOS_1 & C_1^1 & C_1^2 & \dots & C_1^m \\ LOS_2 & C_2^1 & C_2^2 & \dots & C_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ LOS_n & C_n^1 & C_n^2 & \dots & C_n^m \end{bmatrix} \quad (1)$$

To assess the performance of the approach against traditional practice, we also define a baseline method that leverages age as a single feature in prediction.

Beyond LOS prediction, the importance ranking of features in the predictive model could assist HCOs to apply and interpret the corresponding results. The random forest regression model enables importance ranking for each feature, which is calculated as extent to which prediction error increases when data for the investigated feature is permuted while all others are held constant [21].

Performance Evaluation

A random forest regression predicts an LOS as a continuous value. We evaluated the performance of this prediction with respect to the actual LOS as follows. Let us assume LOS_i and \widehat{LOS}_i ($i = 1, \dots, n$) are the true and predicted LOS value, respectively. We calculate the difference between \widehat{LOS}_i and LOS_i as t_i . If t_i is smaller than a predefined tolerance threshold τ , we claim a correct prediction for the i^{th} patient:

$$b_i = \begin{cases} 1, & |\widehat{LOS}_i - LOS_i| \leq \tau \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

The accuracy (Acc) of prediction is thus assessed as

$$Acc = \sum_{i=1}^n b_i / n \quad (3)$$

Experiment Design

This section begins with a description of four comparative models in our framework. We then describe how parameters

are selected and compare the models on a range of values for the parameters. To calibrate the parameters and compare the models, we use 5 randomized runs of 3-fold cross-validation. Finally, we describe a hypothesis testing strategy to ascertain which features significantly influence the LOS prediction.

Models

Table 2 – Comparative Models

Model	Description
M_0	Age
M_1	One year of ICD-9 codes
M_2	Two years of ICD-9 codes
M_3	Three years of ICD-9 codes

Table 2 summarizes the four models. The baseline model M_0 uses age as a lone feature. The other three predictive models (M_1 to M_3) rely on one, two, and three-years worth of historical ICD-9 codes.

Parameter Selection

There are three parameters in the framework that need to be tuned: 1) number of trees in the random forest, 2) τ and 3) number of ICD-9 codes in the model. Since models M_1 , M_2 and M_3 are the same in terms of the prediction algorithm, we leverage M_3 as a representative model to select an optimal value for each of the parameters.

Number of Trees

It has been shown that the performance of random forest regression is relatively insensitive to the number of trees [21]. However, selecting too small of a value can lead to poor accuracy, while too large of a value can lead to excessive computational load. Thus, we evaluated the models over a range for the number of trees. We leveraged the distribution of predictive performance of the resulting random forest to select an optimal number. Specifically, we choose the value that maximizes accuracy and minimizes the number of trees.

LOS Threshold τ

The threshold τ introduced above represents the difference tolerance between the predicted LOS and the true LOS, but it will vary from one HCO to another. Thus, the accuracy in LOS prediction is evaluated under a set of thresholds {5, 12, 24, 36, 48}.

Number of Predictors

The number of ICD-9 codes (i.e., features) in this study is dimensionally large (i.e., 1849 in total). As such, we perform dimensionality reduction to derive a more manageable model. This is accomplished as follows. First, we sort the predictors on their importance in descending order. Second, we select a subset of the features and predict LOS. We choose the subset with the smallest size and highest accuracy.

Model Evaluation

There are three variables in our study and they are the number of trees, LOS threshold, and percent of predictors. Thus, three strategies are designed to evaluate our models in Table 3. In each strategy, two variables are held as constant while the third variable is varied.

Table 3 – Model evaluation strategies

Strategy	# Trees	τ	% Predictors
A	Vary	Constant	Constant
B	Constant	Vary	Constant
C	Constant	Constant	Vary

Feature Discrimination Analysis

A two-sample t-test is applied to compare LOS between patients with and without a certain ICD-9 code. We conduct

this analysis for the ten ICD-9 codes with the highest importance, as derived from the random forest regression models. Specifically, for each investigated ICD-9 code, we test the significance of the LOS difference for patients with and without the code.

Results

Model Paramaterization

Figure 3 depicts the predictive performance of the models on a varying number of trees. The performance was evaluated while varying τ over 5, 12, 24, 36 and 48 hours. We selected 50 trees for our random forests because accuracy grows monotonically up to this point, after which it is constant.

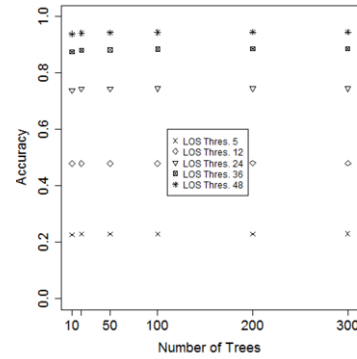


Figure 3 – The accuracy of models as a function of the number of trees in the random forest.

The selection of τ depends on the requirements of an HCO. It can be seen in Figure 3 that the accuracy grows with τ . If an HCO can accept the predicted LOS between a range of 12 hours within the actual LOS, then 12 hours could be selected as the value of τ . Thus, for our investigation, we set τ to be 12 hours.

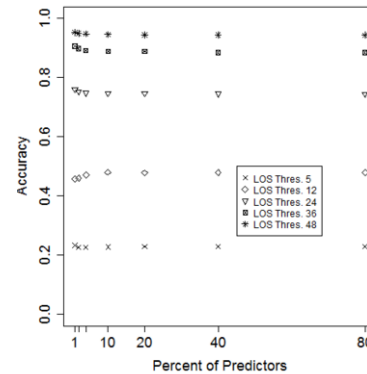


Figure 4 – Model accuracy as a function of the number of features retained.

Our models are based on the top 10% of the features. This is based on the calibration shown in Figure 4, where it can be seen that model accuracy after this point remains relatively constant.

Model Evaluation

Figures 5-7 depict model performance as a function of the number of trees, LOS threshold and percent of features, respectively. In general, it can be seen that the models that incorporate ICD-9 codes have better performance than the

baseline models. Additionally, models M_1 , M_2 and M_3 have almost the same predictive performance. This suggests that one-year of historical ICD-9 codes may be sufficient for LOS prediction.

The accuracy of the four models (where τ is set to 12 hours and the feature set is fixed to the top 10%) on a varying number of trees (10, 50, 100, 200 and 300) is shown in Figure 5. It can be seen that the accuracy of M_1 , M_2 and M_3 is substantially higher than M_0 . Additionally, the number of trees has minimal influence on the accuracy for all of the models.

In Figure 6, it can be seen that as the LOS threshold increases, all models improve in accuracy. However, the predictive performance does not improve when varying the percent of features, as is apparent in Figure 7.

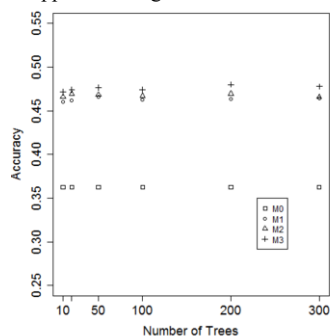


Figure 5 – Model accuracy as a function of the number of decision trees. ($\tau = 12$ hours; predictor set = top 10%)

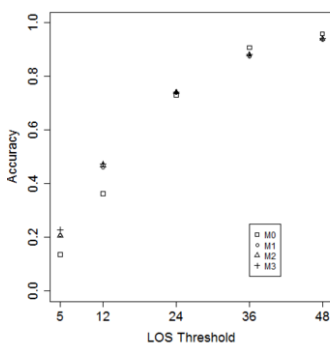


Figure 6 – Model accuracy as a function of the LOS threshold. (Number of trees = 50; Predictor set = top 10%).

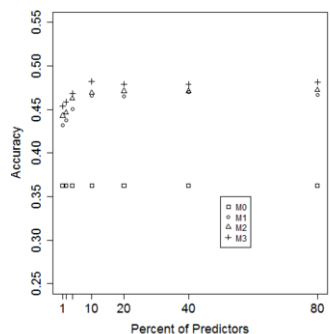


Figure 7 – Model accuracy as a function of percent of predictors. ($\tau = 12$ hours; Number of trees = 50).

Finally, we conducted a series of experiments to predict LOS in two settings. In the first setting, we considered all patients

whose $LOS > 0$. This was done to assess model performance on patients with a range of LOS. In the second setting, we restricted our analysis to the subset of patients whose $LOS \geq 96$ hours. This allows us to study the model for predicting excessively long and costly LOS. To perform this analysis, we compare M_3 with its optimal parameterization (50 decision trees, 12 hours of LOS threshold, and top 10% features) with the baseline model. The results are shown in Table 4. It can be seen that M_3 (49.3%) outperforms M_0 (37.7%) with almost 30% improvement. Moreover, for patients with $LOS \geq 96$ hours, M_3 (5.8%) outperforms M_0 , which is incapable of predicting such excessively long stays.

Table 4 – Model comparison for all patients and patients with $LOS \geq 96$

Model	Accuracy	
	All patients (n = 7683)	$LOS \geq 96$ hours (n = 1505)
M_0	37.7%	0.00%
M_3	49.3%	5.8%

Feature Discrimination Analysis

Out of 1849 ICD-9 codes, 10 were selected for further analysis in terms of their clinical implication (as shown in Table 5). Each of the codes in the top 10 ranks exhibited a statistically significant influence on the LOS between LOS for patients with and without such codes. Such evidence may help healthcare organizations (HCOs) to adopt resource allocation strategies that optimize the care management and improve care quality.

As an example, the most discriminant ICD-9 code was 650: “Normal Delivery”. The mean LOS for patients with and without Normal Delivery in their history was 58.1 and 76.5 hours, respectively. The p -value for this difference was less than 0.001. As another example, patients with ICD-9 code of 659.63 “Elderly multigravida with antepartum condition or complication” stayed about 10 more hours than those without that code (81 versus 70.6).

Discussion

The results show that the historical information in EMRs may assist in forecasting obstetric patients’ LOS in a hospital. Specially, our random forest regression model predicted LOS with an accuracy of 49% under an error range of 12 hours, which is 30% more accurate than a baseline model. The experimental results further demonstrate that a model based on the top 10% of ICD-9 codes can achieve an accuracy as high as those based on all involved ICD-9 codes (over 1800). Additionally, the models based on the most recent year of data logged in the EMRs can achieve similar performance to those based on two-, and three-year worth of data.

Notably, we also investigated the top 10 ICD-9 codes, which have significant differences in terms of LOS between patients with and without such codes. These results suggest that the HCOs can specialize resource allocation strategies accordingly.

Despite the merits of this investigation, we acknowledge that this is a pilot study and there are several limitations. First, the data was collected within a single institution and may not cover all of a patient’s medical history or be readily applicable to another hospital setting. Second, all of the patients in this study were on an obstetric service, such that the framework may not be directly extended to other types of patients or healthcare services. Third, the prediction of LOS may be considered lower than what one might want in a decision support system (i.e., an accuracy of 48% for ± 12 hour). Other

factors that potentially influence LOS can be incorporated in the model, such as certain patient demographics (e.g., race) or

physical traits (e.g., height, weight or BMI).

Table 5 – Summary for the top 10 most predictive ICD-9 codes

Importance Rank	ICD-9 code	Description	# of patients (with code vs. without code)	Mean LOS (with code vs. without code)	p-value
1	650	Normal delivery	2281 vs. 6907	58.1 vs. 76.5	<0.001
2	659.63	Elderly multigravida with antepartum condition or complication	1153 vs. 8035	81.0 vs. 70.6	<0.001
3	V28.81	Encounter for fetal anatomic survey	1899 vs. 7298	75.3 vs. 71.1	0.001
4	285.9	Unspecified anemia	343 vs. 8845	80.5 vs. 71.6	0.044
5	V28.89	Other specified antenatal screening	1580 vs. 7608	73.7 vs. 71.6	0.043
6	V28.4	Antenatal screening for fetal growth retardation using ultrasonics	384 vs. 8804	85.5 vs. 71.3	<0.001
7	V23.9	Unspecified high-risk pregnancy	531 vs. 8657	89.3 vs. 70.9	<0.001
8	V23.82	Supervision of high-risk pregnancy of elderly multigravida	137 vs. 9051	85.4 vs. 71.7	0.035
9	642.93	Unspecified hypertension antepartum	89 vs. 9099	109.8 vs. 71.6	<0.001
10	V76.2	Screening for malignant neoplasm of the cervix	582 vs. 8606	76.5 vs. 71.6	0.008

Conclusions

This paper assessed the feasibility of a machine learning-based framework for predicting the length of stay for obstetric patients using historical diagnoses. We showed that one years worth of diagnostic history may be sufficient to predict hospitalization LOS with accuracy that is substantially higher than a baseline based solely on the age of the patient. We believe this research can be extended by including additional types of historical data (e.g., medications) and leveraging the chronological order of such knowledge.

Acknowledgements

This research was sponsored in part by NIH grants R01LM010207 and R01LM011933.

References

- [1] D. Dranove, C. Garthwaite, B. Li, and C. Ody. Investment subsidies and the adoption of electronic medical records in hospitals. *J Health Econ* **44** (2015), 309-319.
- [2] Y. Chen, J. Ghosh, C. Bejan, C. Gunter, A. Kho, D. Liebovitz, J. Sun, J. Denny, and B. Malin. Building bridges across electronic health record systems through inferred phenotypic topics. *J Biomed Inform* **55** (2015), 482-493.
- [3] C. Yan, Y. Chen, B. Li, D. Liebovitz, and B. Malin. Learning clinical workflows to identify subgroups of heart failure patients. *AMIA Annu Symp Proc* (2016), 1248-1257.
- [4] T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, G. Yang, and Y. Chen. A machine learning-based framework to identify type 2 diabetes through electronic health records. *Int J Med Inform* **97** (2017), 120-127.
- [5] Y. Chen, W. Xie, C.A. Gunter, D. Liebovitz, S. Mehrotra, H. Zhang, B. Malin. Inferring clinical workflow efficiency via electronic medical record utilization. *AMIA Annu Symp Proc* (2015) 416-425.
- [6] Y. Chen, N. Lorenzi, W. Sandberg, K. Wolgast, B. Malin. Identifying Collaborative Care Teams through Electron-ic Medical Record Utilization Patterns. *J Am Med Inform Assoc* **24** (2017) 111-120.
- [7] Y. Chen, S. Nyemba, B. Malin. Auditing medical record accesses via healthcare interaction networks. *AMIA Annu Symp Proc* (2012) 93-102.
- [8] Y. Chen, S. Nyemba, B. Malin. Detecting anomalous insiders in collaborative information systems. *IEEE Trans Dependable Secure Comput* **9** (2012) 332-344.
- [9] WHO, UNICEF, UNFPA. The World Bank, and the United Nations Population Division. Trends in maternal mortality: 1990 to 2015. Geneva, *World Health Organization* (2015).
- [10] G.L. Darmstadt, Z.A. Bhutta, et. al. Evidence-based, cost-effective interventions: how many newborn babies can we save? *Lancet* **356** (2005) 977-988.
- [11] S.K. Schmitt, L. Sneed, C.S. Phibbs. Costs of Newborn Care in California: A Population-Based Study. *Pediatrics* **117** (2006) 154-160.

- [12] K.J. Kerber, J.E. de Graff-Johnson, Z.A. Bhutta, P. Okong, A. Starrs, and J.E. Lawn. 2007. Continuum of care for maternal, newborn, and child health: from slogan to service delivery. *Lancet* **370** (2007), 1358-1369.
- [13] M. Turner, H. Winefield, A. Chur-Hansen. The emotional experiences and supports for parents with babies in a neonatal nursery. *Adv Neonatal Care* **13** (2013) 438-446.
- [14] I.W.M. Verburg, N.F. de Keizer, E. de Jonge, N. Peek. Comparison of regression methods for modeling intensive care length of stay. *PLoS One* **9** (2014) e109684.
- [15] M.G. Goldfarb, M.C. Hornbrook, C.S. Higgins. Determinants of hospital use: a cross-diagnostic analysis. *Medical Care* (1983) 48-66.
- [16] A.M. Epstein, R.S. Stern, J.S. Weissman. Do the poor cost more? A multihospital study of patients' socioeconomic status and use of hospital resources. *N Engl J Med* **322** (1990) 1122-1128.
- [17] L.R. Burns, D.R. Wholey. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Medical care* **29** (1991) 251-271.
- [18] E.J. Federman, C.E. Drebing, C. Boisvert, W. Penk, G. Binus, R. Rosenheck. Relationship between climate and psychiatric inpatient length of stay in Veterans Health Administration hospitals. *Am J of Psychiatry* **157** (2000) 1669-1673.
- [19] M. Lindholm, J.L. Hargraves, W.J. Ferguson, G. Reed. Professional language interpretation and inpatient length of stay and readmission rates. *J Gen Intern Med* **27** (2012) 1294-1299.
- [20] R.A. Deyo, D.C. Cherkin, M.A. Ciol. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* **45** (1992) 613-619.
- [21] A. Liaw, M. Wiener. Classification and regression by randomForest. *R news* **2** (2002) 18-22.

Address for correspondence

Cheng Gao: cheng.gao@vanderbilt.edu

Group-Based Trajectory Analysis for Long-Term Use of Warfarin Therapy in Atrial Fibrillation Patients

Shijing Guo^a, Xiang Li^a, Xin Du^b, Haifeng Liu^a, Guotong Xie^a

^a IBM Research - China, Beijing, China,

^b Department of Cardiology, Beijing Anzhen Hospital, Beijing, China

Abstract

Atrial fibrillation (AF) patients suffer a high risk of ischemic stroke and other thromboembolism (TE). Warfarin is a long-term oral medication and is effective in reducing TE for AF patients. Identifying the trajectory patterns of warfarin use in AF patients and discovering how different trajectories are associated with different TE outcomes are important for understanding long-term use of warfarin. Also, finding the factors affecting future warfarin use and predicting the warfarin use trajectory for new patients can help to efficiently target the specific patient groups and propose relevant interventions for warfarin use. This paper, combining group-based trajectory modeling and predictive modeling, has successfully discovered three patient groups with distinct warfarin use trajectories. Also, results suggest that the warfarin use trajectory has potential association with the TE outcomes. Moreover, results show factors affecting the future trajectory, which have been used to build prediction models for the warfarin use group.

Keywords:

Anticoagulants; Atrial Fibrillation; Stroke.

Introduction

Atrial fibrillation, also called AFib or AF, which is a common cardiac rhythm disturbance, can greatly increase the risk of ischemic stroke and other thromboembolism (TE) in patients [1]. Warfarin is highly effective in reducing TE in AF patients [2], and is commonly prescribed as a long-term oral anticoagulant for AF patients. However, relevant studies [3] have shown that patients have substantial difficulties in maintaining adherence to warfarin regimens. Thus, it is important to study warfarin adherence and its outcomes for AF patients.

Many researchers have discovered the influence of patient adherence to warfarin on stroke prevention [4] and the factors associated with non-adherence to warfarin therapy [5]. However, how warfarin adherence changes over a period of time in an AF patient group and its relevant impact on TE outcomes still remain unclear, and patients' future trajectory prediction of warfarin use needs to be further studied. Identifying the underlying trajectory patterns of long-term warfarin use can provide deeper understanding of how different trajectories are associated with different TE outcomes. Moreover, discovering the factors that affect the future trajectory and predicting the trajectory group for new patients can help to provide personalized interventions of warfarin use for specific patient groups. For example, it could be predicted that the warfarin use trajectory of a patient would

change at a certain time, and relevant interventions could aim for this specific time point.

This paper, continuing our previous work on AF studies [6-8], focuses on the warfarin use over a two-year time period from Chinese Atrial Fibrillation Registry (CAFR) data [9]. It groups patients according to similar trajectories of warfarin use, and discovers the underlying trajectory patterns of the warfarin use for each group. Also, the impact of different trajectories on TE outcomes is observed. Furthermore, feature selection is conducted in order to find group characteristics and to identify the factors associated with the future trajectory of warfarin use. Additionally, prediction models built for new patients are able to predict the trajectory group membership using patient baseline information.

Methods

The study pipeline is illustrated in

Figure 1. It integrates group-based trajectory modeling, feature selection and prediction modeling.

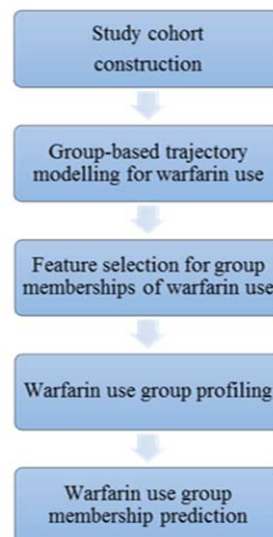


Figure 1— Pipeline of Group-based Trajectory Study for Warfarin Use

The pipeline includes five steps:

- study cohort construction, which is to identify the study patient group;
- group-based trajectory modeling, which aims to cluster patients according to similar trajectories of warfarin use;
- feature selection, which selects key features associated with the warfarin use trajectory;
- warfarin use group profiling, which is to discover patient characteristics in different trajectory groups;
- warfarin use group membership prediction, which is to predict the future group of warfarin trajectory for new patients.

Methods used in each step are explained in detail in the following subsections.

Cohort construction

The first step constructs the study cohort. The study cohort data are from Chinese Atrial Fibrillation Registry (CAFR) data [9], which covers over 17,000 AF patients from 32 hospitals in Beijing, China. The CAFR dataset was collected from 2011 to 2015, and contains both baseline and follow-up data. The baseline data contains patient demographics, signs, symptoms, medical history, physical examination and laboratory test results, and current treatments at the time of the registry. The follow-up data record patient treatments and clinical events such as TE.

In order to observe the use of warfarin over a period of two years and exclude the impact of radiofrequency ablation (RFA), which also shows significant effectiveness in reducing TE occurrence, AF patients who meet the following criteria are selected as the patient cohort for the study:

1. Patients used warfarin at baseline
2. Patients did not receive radiofrequency ablation (RFA) from baseline to the 24-month follow-up.
3. Patients have a complete two-year follow-up record of warfarin use after data imputation.

A total of 1348 patients who meet our criteria are selected as the study cohort.

Figure 2 illustrates the timeline of the study. The warfarin use observation time is a two-year period, including the time points at baseline, the 3rd, 6th, 12th, 18th and 24th month follow-ups. In particular, data at the first time point indicates whether warfarin is used at baseline, data at the remaining time points indicate whether warfarin is used between the last follow-up time and current follow-up time. Similarly, the TE occurrence rate is observed over a two-year period and the period includes four data-collection time points. The data at each time point mean whether a TE occurred in the last six months.

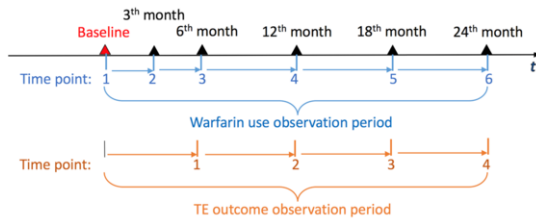


Figure 2– Study Timeframe for Warfarin Use

Group-based trajectory modeling

The second step is to cluster patients by their observed data of warfarin use and to group patients with similar trajectories together. Group-based trajectory modeling is conducted at this step.

Group-based trajectory modeling has been increasingly applied in clinical research [10-12], such as for mapping the developmental course of symptoms and assessing heterogeneity. The group-based trajectory model adopted from Nagin, 2005 [13] assumes that the appropriate parametric model $f(y)$ to be studied can be presented as a mixture model, as below:

$$1) f(y) = \sum_{k=1}^K \Pr(C = k) \Pr(Y = y|C = k)$$

where $y = (y_1, y_2, \dots, y_T)$ denotes the longitudinal sequence of an individual's behavioral measurements, which are the measurements of warfarin use in this study, over the T periods, C is the given group membership, and $k = (1, 2, \dots, K)$ is this individual's group membership [14].

Since the warfarin measurement data is binary, the following likelihood [14] of observing the trajectory for patient i, given that the patient belongs to group k, is:

$$2) \Pr(Y_i = y_i | C_i = k, W_i = w_i) = \prod_{y_{ij}} p_{ijk} \prod_{y_{ij}=0} (1 - p_{ijk})$$

$$\text{with } p_{ijk} = \frac{\exp(\sum_{n=0}^N \beta_{nk} t_{ij}^n)}{1 + \exp(\sum_{n=0}^N \beta_{nk} t_{ij}^n)}$$

where w is the time dependent variable t (i.e. 'month') in this study, j is the index of the time point, and n is the order of the polynomial equation.

In order to determine the relevant parameters, expectation maximization (EM) is used to find the max likelihood of the mixture model. Bayesian information criterion (BIC) which considers the likelihood, the sample size and number of parameters is used for the selection of the number of the group, 'K'.

At the end of this step, patients with similar trajectories of warfarin use are grouped together.

Feature selection

In order to discover which baseline features affect the future trajectory of warfarin use, factor selection for group memberships are performed at this step.

In this study, the target is the group membership which is a multi-categorical variable. This step implements filter-based feature selection methods, and measures one risk feature at a time about how important each feature predicts the group membership. The metric is the p-value which indicates the significance of the association between the feature and the target. Specifically, the dataset of this study includes two types of features: categorical features, such as whether the patient has hypertension, and continuous features, such as age.

- For categorical features, the p-value based on Pearson's Chi-square is implemented as the filter-based feature selection method.
- For continuous features, the p-value based on the ANOVA F Statistic is applied.

Group profiling

Based on the features selected from the last step, characteristics of each trajectory group discovered by the group-based trajectory modeling are summarized in this step in order to discover the distinctions between different trajectory groups. Moreover, TE occurrence rates of different

trajectory group are observed, which aims to show how different warfarin use trajectories may lead to different TE risks. Whether a TE occurs in the last six months at each time point are compared.

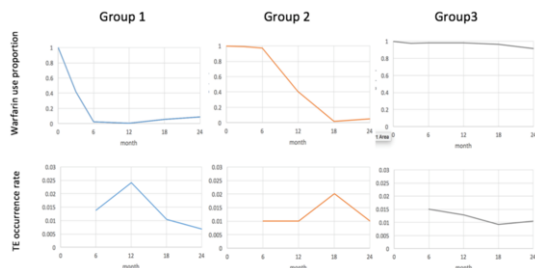


Figure 3— Fitted trajectories of warfarin use for AF patients and the corresponding TE occurrence rates

Group membership prediction modeling

This step is to predict the trajectory group based on baseline features. During this process, three feature sets are applied and compared, including:

- all baseline features, which are 121 in total and cover patient demographics, signs, symptoms, medical history, physical examination and laboratory test results, medication and treatment information.
- CHA2DS2-VASc factors [15], which include age, prior TE, congestive heart failure, hypertension, diabetes mellitus, vascular disease and sex.
- Selected features using the filter-based feature selection method.

Since the prediction model aims to predict the group membership of new patients based on the baseline features, we choose to conduct and compare three interpretable prediction modeling methods in this study, which are listed as below.

- Multinomial logistic regression modeling: Since the target is a multi-categorical variable, the multinomial logistic regression modeling is implemented.
- Binary logistic regression modeling: During this modeling process, each individual group is predicted respectively with the remaining groups, and the target is transformed into a binary variable in this modeling process.
- Decision tree modeling: Chi-squared Automatic Interaction Detector (CHAID) algorithm is applied to build a tree-based prediction model, due to its good interpretability.

Each model is implemented using the above three feature sets.

During the predictive modeling process, input data is randomly divided into two datasets which are used for model training and model testing respectively. The ratio of the dataset for model training to the dataset for model testing is 6:4.

In order to avoid the impact of highly imbalanced data, models are evaluated using the metric *the area under the receiver operating characteristic curve* (AUC) instead of the accuracy of the prediction.

Results

Group-based trajectories of warfarin use in AF patients

Group-based trajectory modeling has successfully discovered underlying longitudinal trajectories of warfarin use among AF patients. According to the similar trajectories, three patient groups are identified using BIC as the model selection criterion. Moreover, the TE occurrence rates are observed and the model shows that TE occurrence rates have potential association with the warfarin use trajectory.

Results are summarized in Figure 3. As we can see, three patient groups have three distinct patterns of warfarin use trajectories. The first patient group has the worst warfarin adherence. The warfarin use proportion in group 1 appears to instantly drop after baseline, and it remains near zero in the following 18 months. The warfarin use in group 2 maintains a high proportion of use in the first six months. Then, warfarin use declines to a near-0 level over a period of one year, and the level of use remains at the same proportion from the 18th month to the 24th month. Patients in group 3 have the best warfarin adherence compared with the other two groups. Most patients continually use warfarin and the warfarin use proportion of group 3 stays close to '1' over the 24-month study period.

The TE occurrence rates are observed to closely associate with warfarin use trajectories, which are illustrated in Figure 3. Both group 1 and group 2 observed an increase in TE occurrence rate after the decline in warfarin use. In particular, group 1 shows a rapid decline in the first six months. The TE occurrence rate appears to significantly increase around the 12th month, and the six-month occurrence rate in these AF patients can be up to 2.4%. Compared to group 1, group 2 has a postponed decline of warfarin use between the 6th and 18th month, and the TE occurrence rate of group 2 observes a corresponding delayed increase. The peak point of the TE occurrence rate in group 2, which is up to 2% at the 18th month, is lower than the peak rate in group 1. Compared with the other two groups, patients in group 3 maintain a high level of warfarin use over two years, and the relevant TE occurrence rate is controlled to under 1.5% in this study period.

Feature selection and group profiling

A total of 29 baseline features which are associated with the patient group membership for warfarin use trajectory are discovered. Based on the discovered features at baseline, the characteristics of each group are summarized in Figure 4.

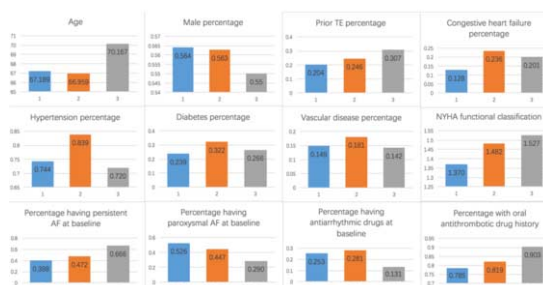


Figure 4— Comparison of the warfarin use group characteristics

In particular, compared with patients from group 1 and group 2, patients from group 3 who maintain a high percentage of warfarin use over the study period have the following significant characteristics:

- an older average age than the other two groups
- highest average score in New York Heart Association (NYHA) functional classification
- highest percentage of patients with TE history
- highest percentage of patients with persistent AF at baseline
- highest percentage of oral antithrombotic drug use history

Prediction for group membership of warfarin use

In the process of predicting group membership of warfarin use, we applied three feature sets in this study: all features, CHA2DS2-VASc features, and selected features using the feature selection method. For each feature set, three types of modeling methods are implemented respectively: multinomial logistic regression modeling, binary logistic regression modeling and decision tree modeling. AUC is used for the evaluation of the prediction models. Each patient group of warfarin use, which is group 1, group 2 or group 3, is predicted separately. The results of AUC for each group and their average AUCs are described in Table 1.

Table 1- AUC of different models on different feature sets

Features	Modeling	AUC on Training Set				AUC on Testing Set			
		Group1	Group2	Group3	Average	Group1	Group2	Group3	Average
All features	MLR ^a	0.869	0.902	0.88	0.884	0.63	0.552	0.67	0.617
	BLR ^b	0.863	0.89	0.857	0.870	0.656	0.558	0.664	0.626
	CHAID ^c	0.817	0.758	0.781	0.785	0.67	0.59	0.647	0.636
CHA2DS2-VASc factors	MLR	0.599	0.664	0.627	0.630	0.619	0.573	0.587	0.593
	BLR	0.599	0.665	0.628	0.631	0.619	0.571	0.59	0.593
	CHAID	0.598	0.65	0.598	0.615	0.618	0.592	0.538	0.583
Selected features	MLR	0.771	0.745	0.762	0.759	0.757	0.616	0.748	0.707
	BLR	0.773	0.74	0.759	0.757	0.756	0.614	0.747	0.706
	CHAID	0.775	0.735	0.77	0.760	0.717	0.596	0.687	0.667

a. Multi-nominal logistic regression
b. Binary logistic regression
c. Decision tree

As we can see, the training dataset has the highest AUC score when all baseline features are used for the prediction. However, applying all possible features can increase the redundancy among features and can cause overfitting problems, which are seen in the AUC scores of the testing dataset. The AUC scores have reduced to 0.617 for group 1, 0.626 for group 2 and 0.636 for group 3 prediction.

CHA2DS2-VASc factors have the lowest average AUC. This indicates that CHA2DS2-VASc factors are not enough in predicting the group membership of warfarin use.

The AUC scores significantly improve when the 29 selected features are applied, which indicates that the selected features have the best performance in predicting the warfarin use trajectory.

Moreover, different modeling method results using the same selected features show that multinomial regression modeling and logistic regression modeling have similar AUC scores. The decision tree model has a lower AUC score than the other two modeling methods for the testing dataset, though it performs well on training set. Specifically, multinomial regression modeling with the selected 29 features have the highest average AUC score as well as the highest AUC scores for each group's prediction in the testing set. The AUC scores are 0.757 for group 1, 0.616 for group 2 and 0.748 for group 3 prediction respectively.

Discussion

This paper has discovered underlying longitudinal trajectories of the warfarin use, and has found the association between the trajectories and the TE occurrence rate. In particular, results

indicate that patients in group 3 are older and have more serious AF conditions than patients in group 1 and group 2, according to the NYHA functional classification, the percentage of patients with TE history, and the percentage of patients with persistent AF at baseline. However, the persistent use of warfarin in group 3 leads to better control of TE occurrence than the other two groups.

The prediction for warfarin use group membership could be applied in many scenarios. For new patients, the prediction of the group membership can help to understand the future trajectory of warfarin use. According to the patterns of the trajectory, such as the time point of the use decline, more personalised interventions could be proposed.

The limitation of this study is that the number of the AF patients who meet the study participant selection criteria is small, because a large percentage of the registered AF patients have received the RFA treatment at some timepoint during the two-year study period. The small dataset and the imbalanced cases limit some of the studies, such as TE risk model for each trajectory group. Future work includes collecting data about the quantities of warfarin use, covering more time points, and conducting more detailed trajectory analysis for larger datasets.

Moreover, this study aims to analyse one variable only, which is the use of warfarin in this study. Future work can extend to multiple variables and conduct multi-trajectory modeling [11], which considers the changes of more than one variable over time.

Conclusions

This study, combining group-based trajectory modeling and predictive modeling, has discovered three distinct trajectories of two-year warfarin use. Patients are classified into three groups, and each group's characteristics have been summarized. Moreover, results find that the trajectories of warfarin use are closely related to TE outcomes. Additionally, 29 baseline features for the trajectory group are identified. By comparing three feature sets and three types of models, it shows that the multinomial modeling with the selected 29 features obtains the best performance in predicting the group membership of the warfarin use trajectory.

References

- [1] C.T. January, L.S. Wann, J.S. Alpert, H. Calkins, J.E. Cigarroa, J.C. Cleveland, et al, Guideline for the management of patients with atrial fibrillation. *Journal of the American College of Cardiology*. **64** (2014) :2246–80.
- [2] E.N. Prystowsky, D. Woodrow Benson, V. Fuster, R.G. Hart, G. Neal Kay, R.J. Myerburg, G.V. Naccarelli, D. George Wyse, Management of Patients with Atrial Fibrillation, *Circulation*. **93** (1996), 1262-1277
- [3] S.E. Kimmel, Z. Chen, M. Price, C.S. Parker, J.P. Metlay, J.D. Christie, C.M. Brensinger, C.W. Newcomb, F.F. Samaha, R. Gross, The influence of patient adherence on anticoagulation control with warfarin: results from the International Normalized Ratio Adherence and Genetics (IN-RANGE) Study. *Arch Intern Med*. **167** (2007), 229-35.
- [4] P.P. Kneeland, M.C. Fang. Current issues in patient adherence and persistence: focus on anticoagulants for the treatment and prevention of thromboembolism. *Patient Prefer Adherence* **4** (2010), 51-60.
- [5] S.V. Pamboukian, I. Nisar, S. Patel, L. Gu, M. McLeod, M.R. Costanzo, A. Heroux. Factors associated with non-adherence to therapy with warfarin in a population of chronic heart failure patients. *Clin Cardiol*. **31** (2008), 30–34.
- [6] X. Li, H. Liu, X. Du, P. Zhang, G. Hu, G. Xie, S. Guo, M. Xu, and X. Xie, Integrated Machine Learning Approaches for Predicting Ischemic Stroke and Thromboembolism in Atrial Fibrillation, in: *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, 2016, p. 799.
- [7] X. Li, H. Liu, X. Du, G. Hu, G. Xie, P. Zhang, Using Frequent Item Set Mining and Feature Selection Methods to Identify Interacted Risk

- Factors - The Atrial Fibrillation Case Study. *Studies in Health Technology and Informatics* **228** (2016), 562-6.
- [8] S. Guo, X. Li, H. Liu, P. Zhang, X. Du, G. Xie, F. Wang, Integrating Temporal Pattern Mining in Ischemic Stroke Prediction and Treatment Pathway Discovery for Atrial Fibrillation. *AIMA Joint Summits*, American Medical Informatics Association, 2017.
 - [9] X. Du, C. Ma, J. Wu, S. Li, M. Ning, R. Tang, et al. Rationale and design of the Chinese Atrial Fibrillation Registry Study. *BMC Cardiovascular Disorders* **16** (2016), 130.
 - [10] D. Nagin, and C. L. Odgers, Group-based Trajectory Modeling in Clinical Research. *Annual Review of Clinical Psychology* **6** (2010), 109-38.
 - [11] P. Burckhardt, D.S. Nagin, and R. Padman, Multi-Trajectory Models of Chronic Kidney Disease Progression, in: *AIMA Annual Symposium Proceedings*, American Medical Informatics Association, 2016, p. 1737.
 - [12] J.M. Franklin, W.H. Shrank, J. Pakes, G. Sanf lix-Gimeno, O.S. Matlin, T.A. Brennan, N.K. Choudhry. Group-based trajectory models: A new approach to classifying and predicting long-term medication adherence. *Medical Care*, **51** (2013), 789-796.
 - [13] D. Nagin, 2005. Group-based Modeling of Development. Cambridge, MA: Harvard University Press
 - [14] B.L. Jones, D.S. Nagin, and K. Roeder, A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories. *Sociological Methods & Research* **29** (2001), 374-393.
 - [15] G.Y. Lip, R. Nieuwlaat, R. Pisters, et al. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest*, **137** (2010), 263-72.

Address for correspondence

Shijing Guo. Address: Building 19, Zhouguancun Software Park 8
Dongbeiwang West Road, Haidian District Beijing, P.R.C. 100094.
E-mail: bjsgjguo@cn.ibm.com.

CTdBem - A New Protocol for Ultra Low Radiation Dose MDCT

Carlos Jesus Pereira Haygert^a, Marcos Cordeiro d'Ornellas^b,
Daniel Welfer^b, Rubens Martins Bastos^a, Gustavo Nogara Dotto^a

^a e-Health and Imaging Diagnostic Unit, University Hospital of Santa Maria, Santa Maria, RS, Brazil

^b Department of Applied Computation, Federal University of Santa Maria, Santa Maria, RS, Brazil

Abstract

This study describes a novel method of obtaining low-dose computed tomography (CT) scans followed by imaging postprocessing that provides diagnostic quality to such low-dose exams. In addition, we compared the Total Radiation Doses (DLP) of the 64-channel MDCT x 16-channel MDCT for a new Dental CT - CTdBem protocol for hospital use. DLP data obtained from 20 patients using 16-channel MDCT was compared with 20 other patients using 64-channel MDCT. In both tomographic (Aquilion 64 and Brightspeed 16) FOV was approximately 160(V) x 130(H) mm. An imaging postprocessing algorithm was used to provide diagnostic quality to the obtained low-dose CT scans. Imaging postprocessing included imaging smoothing, multiplanar reconstruction (MPR), and volume rendering (VR), as well as surface rendering (SR) to allow three-dimensional printing of the desired scans. The average DLPs were of 28,5 mGy.cm and 54,65 mGy.cm, using the 64-channel MDCT and 16-channel MDCT, respectively. The effective radiation dose (DLP) from the 64-channel MDCT statistically differs from the DLP data from 16-channel MDCT, resulting in a value of $p < 0.05$ for all comparisons. A novel low-dose CT protocol for dentomaxillofacial assessment using imaging postprocessing techniques is described. The authors concluded that although the DLP values differ statistically ($p < 0.05$), both equipment (64 and 16-channel MDCT) produce tomographic images of patients with low radiation doses. The greater the number of detectors, the lower the mAs product and, consequently, the amount of X-radiation produced.

Keywords:

Radiography, Panoramic; Cone-Beam Computed Tomography; Multidetector Computed Tomography

Introduction

Currently, there is a growing concern in society regarding the radiation doses to which we are exposed while conducting tests using ionizing radiation. Such concern is not only associated to patients but also to health professionals and equipment manufacturers [1]. In this context, CT technology has advanced over the years, as well as the increase of its use and applications, that have brought deep concerns regarding the radiation [2]. Moreover, there is the outstanding fact that the medical imaging modality with the highest dose of radiation is obtained from the CT. From this, a reduction of this dose has become a very important goal in these applications [3].

The Multislice Computed Tomography (MDCT) has a spiral/helical scanner with more than one row of detectors, having 4, 16, 64, 256, 320 or even 520 lines (channels) of detectors, which are able to generate many slices simultaneously and several complete analyzes in seconds or in a period of a subsecond.

For dental procedures, both clinical and hospital, the use of imaging tests to analyze bone structure and teeth is considered an important aid to the diagnosis. In this regard, in hospital procedures, the use of MDCT for diagnosis and dentomaxillofacial evaluation of patients becomes a great ally, since this is a commonly found device in hospitals and, when properly calibrated, enables the acquisition of images displaying bone and tooth structures in the region.

The aid in diagnosis that this imaging method provides on craniofacial structure of patients enables the evaluation of bone and tooth structures using multiplanar slices (axial, sagittal, coronal, panoramic and transversal) and reconstruction in 3DVR (three dimensions – rendering volume). Medical tomography may be, then, integrated to dentistry, when kVp (kilovoltage), mAs (milliamperage) and the FOV (Field of View) of the equipment are modified and fixed, which also provides shorter duration to tests [4]. The FOV considers the size of the desired field of view for the exam and has standard sizes for detectors for certain types of examination, considering the appropriate size of each display area [5].

Despite that, imaging evaluation of teeth and facial jaws of patients hospitalized at the University Hospital of Santa Maria (HUSM) did not occur prior to radio and chemotherapy treatments, even given the fact that prior to these procedures the need for dental treatment must be checked, especially when those are related to face, since the follow up time for new dental procedures after radiotherapy and/or chemotherapy with no risk for the patients is usually five years.

However, after discovering an available MDCT equipment at HUSM, the founders of this project started using the MDCT with reduced radiation dose protocol for dentomaxillofacial evaluation in the hospital and reformatted DICOM images acquired in the MDCT in axial, sagittal, coronal, panoramic, transversal cuts and 3D reconstructions, as well as a report of all tests. Thus, the planning and possibilities for care and treatments for patients at HUSM are expanded.

This study aims to compare the Total Radiation Doses (DLP) of 64-channel MDCT with radiation doses of 16-channel MDCT for a new Dental CT - CTdBem protocol for hospital use.

Methods

The following CT scanners were used for the tests: a) Aquilion 64 (Toshiba America Medical Systems, Inc., Tustin, CA, USA) using 120 kV, 10mAs and acquisition time of about 6 seconds (0.5s to 360°/rotation); and b) GE BrightSpeed 16 (GE Medical Systems, Waukesha, WI, USA), using 120 kV, 10mAs and acquisition time of about 12 seconds (0.5s for 360° rotation). In both tomographic (Aquilion 64 and Brightspeed 16) FOV was approximately 160(V) x 130

(H) mm. Multislice CT scans were performed with patients lying in supine, Frankfurt plane perpendicular to the ground, sagittal plane perpendicular to the ground and occluded teeth.

After CT image acquisition, an imaging postprocessing algorithm using open source software (Horos™, Horos Project) was applied to all exams. It involved image smoothing aiming to reduce image noise and provide diagnostic quality for low-dose tomographic images. Subsequently, multiplanar reconstruction (MPR) and three-dimensional reconstruction using volume rendering (VR) was carried out in order to obtain sagittal, coronal, and 3D images of the bucco maxillofacial region. Surface rendering (SR) reconstructions (also performed using Horos™) provided .stl files that enabled preoperative 3D printing of the surgeon's region of interest.

Statistical analysis was performed from the CT scanner's console, in which it is possible to obtain estimated dose data: DLP (Dose Length Product), referent to the total dose of radiation produced during the acquisition. In this research, data from independent K-samples, a single variable (DLP - radiation dose) of numerical data, were used. In order to statistically test the data, the paired T-Test was used aiming to verify the differences between the means. Dose data from 20 (twenty) procedures performed in MDCT 64 were used for comparison with dose data from 20 (twenty) procedures performed in MDCT 16.

Because it is an academic research project in which the focus is to reduce the radiation dose in MDCT at a university hospital, there are no criteria for inclusion or exclusion of patients, given that the work was performed in an environment that aims to serve all those who need its services. However, all imaging tests were only conducted after clinical evaluation and appropriate referral in hospital forms, reporting the necessity of the procedure.

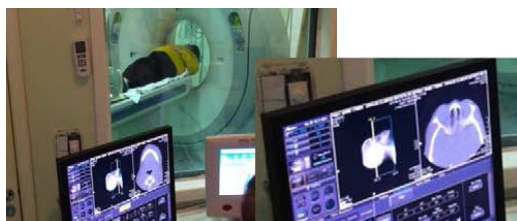


Figure 1 – Control room and Multi Detector CT scanner monitors during the acquisition of tests. One can observe the patient lying on the table, head inside the gantry, scout and sequence of cuts being acquired on the computer screen



Figure 2 – Computer program screen with examination of a patient in Dental CT format. One can observe the axial, panoramic and transversal cuts beside the 3DVR reconstruction

Results

According to the data acquired using 64-channel MDCT with 120 kV and 10mAs protocol during exposure, we obtained an average of 28,5 mGy.cm for DLP and a total scan time of approximately 6 seconds for the entire face. For the 16-channel MDCT with 120 kV and 10mAs protocol during the exposure, we obtained an average of 54,65 mGy.cm for DLP and a total scan time of 12 seconds for the entire face. Through the analysis of Table 1 it is possible to observe the difference in mean values of total radiation dose (DLP) obtained in different methods. The statistic software used was Minitab version 17 for Windows (Minitab Inc. State College, Pennsylvania, USA).

Table 1 – Paired T-test with confidence interval (CI) for the comparison of total radiation dose (DLP) means obtained in the tests performed in Aquilion 64 and GE BrightSpeed 16 MDCTs. The effective radiation dose (DLP) for 64-channel MDCT statistically differs from the DLP data from 16-channel MDCT, resulting in a value of $p < 0.05$ for all comparisons.

Table 1 – Paired T-test and CI: Aquilion 64, Brightspeed 16

Group	N	Mean	Std. Dev	Std. Error
Aquilion 64	20	28.475	3.060	0.684
Brightspeed 16	20	54.650	4.309	0.964
Difference	20	-26.18	4.84	1.08

Note: 95% CI for mean difference (-28.44; -23.91); T-test for mean difference = 0 (vs $\neq 0$); T-value = -24.20; p-value = 0.000

Discussion

When it comes to performing a Computed Tomography (CT), the radiation dose the patient receives depends on the device design and technical parameters and it can vary from one machine to another in different types of acquisition. The use of medical tomography (MDCT) integrated to dentistry allows the acquisition of good quality images with low radiation exposure, since standard protocols parameters for equipment use, such as kilovoltage (kVp) and electric current (mAs), can be modified. [4]. Furthermore, the use of correct FOV for the desired field of view also interferes in the process [5].

Given that, we highlight that in dental procedures, both clinical and hospital ones, what it is important to evaluate are bone and tooth structures and, from that, a protocol that

evidence these tissues without compromising image and optimizing the radiation dose during test acquisition can be determined.

A comprehensive approach to radiation dose management must be focused on providing the right quantity and quality of radiation, as and when is required. Most Computed Tomography scanners (CT) are programmed to radiation dose parameters and automatically archived in correspondence to the scanning parameters defined for each exposure protocol (6). Still, radiation dose reduction must be a priority when it comes to CT, ie, the highest spatial resolution is not a priority on the need to optimize the radiation dose [6].

As new technologies for CT have been developed, a dose reduction of up to 80% can be claimed while maintaining image quality of the diagnosis. In this context, advances in tomographic image are designed to provide equivalent diagnostic image quality, with similar appearance to the full dose images, at a fraction of the radiation dose. Those technologies aim to automatically adjust the equipment to the patient, compensating parameters for the particular physiology and the optimization of dose for desired anatomical region [2].

The energy of the X-ray beam (kilovoltage), the tube current (milliamperage), the rotation time (equal to the exposure time), slice thickness (collimation), pitch (defined as the distance on a 360° rotation table divided by slice thickness), distance of the X-ray tube and the scanning length [7] are included as some of the parameters that can be manipulated and have direct influence on the radiation dose. With the reduction of kVp and mAs for dose optimization in routine tests on CT it was found that there are no qualitative and quantitative losses in image for diagnosis. In this case, the reduction of radiation dose was justified using techniques that consider individual anthropometric data or with less conservative noise levels [8]. With the same effective radiation dose, approximately 44%, in pulmonary CT angiography from the kilovoltage (kVp) reduction and increase in noise levels (35). In accordance to that, tests with four different levels of electrical current (150, 100, 70 and 40mA) were carried out when performing a Volumetric Helical Thorax CT. They have found that the reduction in current did not generate significant difference in image quality [9].

The performance of CT imaging systems is influenced by the specific tomography system, concerning the current (milliamperage - second), kilovoltage (kVp), slice thickness, pitch size and collimation of X-ray beam, as well as image processing and its visualization. Therefore, these factors must be adjusted to optimize image quality when it comes to performance of low image contrast (LCD - low contrast detail), reducing image noise and maintaining lower radiation dose to patients [3].

The relation between the energy of the X-ray beam (kVp) and the radiation dose is not linear, while the one between the tube current and the radiation dose is, which means that the increase of the tube current in 50% will result in a dose 50% higher. Tube current (mA) and exposure time (in seconds) are associated (milliamperage per second - mAs), also leading to a linear relation with the resulting radiation dose [7].

In this context, the acceptable radiation dose is determined by the clinical situation. When decreasing the mAs, for example, the radiation dose is reduced, but, simultaneously, the image noise is increased and the contrast - noise ratio (CNR) is reduced. However, the acceptable compensation level in image quality must be determined in accordance to the purpose of clinical diagnosis and to the task to be performed.

Low kilovoltage increases photoelectric interactions, which improves the level of attenuation and leads to an increase in image contrast and better visualization of details [3]. In this setting, we highlight the important role that our postprocessing algorithm played in the low- dose CT protocol: as we obtained lower radiation dose by decreasing tube current (measured in mAs), we consequently dealt with higher image noise if compared with conventional multislice computed tomography scans. By reducing noise through image smoothing, as well as enabling multiplanar reconstructions, volume rendering, and surface rendering, our open source-based postprocessing algorithm increased the image quality of the low-dose CT scans, allowing their use to diagnose and plan procedures, including three-dimensional printing, with significantly lower radiation doses than conventional CT scans.

Although the image noise measured is higher in low kVp images, the subjective quality of image is higher for lower kVp than for larger ones, since they do not compromise the diagnosis [10]. Still, it is emphasized that the radiation dose to patients is not linear with the kilovoltage. However, reducing kVp can diminish the amount of radiation when other exposure factors are fixed. Furthermore, although the selection of low kilovoltage minimizes radiation for patients, kVp must be selected depending on the diameter of the transversal section of the patient and adjusted according to the task [7].

Thus, it is clear that the interdependence between image quality and radiation dose related to kVp is quite complex. The kVp must be optimized to be low enough to increase the contrast resolution, in order to improve LCD (low contrast detail), but high enough to reduce noise and minimize radiation [3].

Due to the variety of clinical experiences and preferences, image quality varies and it is difficult to determine a fixed set of image acquisition parameters that provides the required image quality. However, that can be achieved based on subjective values. For these reasons, local protocols are developed based on local experience (29). It is important to consider that, like every process at an institution, CT protocols should be regularly reviewed to ensure that image quality and radiation dose are really optimized. It is noteworthy that overdosing and underdosing are medical errors and thus, dose reduction and increase may be appropriate or not when trying to optimize CT protocols [7].

The authors of this paper sought to validate the radiation dose optimization of MDCT and its use in dental diagnosis in hospitals by comparing the DLP obtained in 64-channel MDCT with the DLP obtained in 16-channel MDCT, both using low radiation protocol. In order to do this, the study used as application the parameters for performing CT Dental - CT do Bem (by the same author), which is held at HUSM with the 16 and 64 MDCT calibrated in 120 kVp and 10mAs, with approximate time acquisition of six (6) seconds to 64- channel MDCT and twelve (12) seconds to 16-channel MDCT using FOV of 160mm x 130mm for both equipment. Although the acquisition area of 64-channel MDCT is larger than the 16-channel MDCT, around 11cm and 5.5cm respectively, the reduced pitch causes the slice area acquired in each rotation/turning to differ in only 2cm between scanners. This explains why the test acquisition time is twice (in seconds) and not 4 times longer as it would be expected.

Since it is an innovative work, the comparison of DLP in different equipment, 16-channel MDCT and 64-channel MDCT with altered protocols for dose reduction and visualization of important images for dental diagnosis, is made

from an image base determined by local/subjective clinical factors that produced significantly positive results in terms of image quality available during the development of CTdBem research project.

The authors still consider relevant to clarify that the MDCT is a machine commonly used in most hospitals and therefore, should be used to provide images for dental evaluation purposes. We emphasize that, even when facing low radiation dose protocols, CT scans should only be performed under proper medical guidance.

Conclusion

The authors concluded that although the DLP values statistically differ ($p < 0.05$), both equipment (64 of patients with low radiation doses). The greater the number of sensors present, the lower the mAs product and consequently, the lower the amount of X radiation produced. Even though lower mAs lead to higher image noise, by using a post-processing imaging algorithm it was possible to use low-dose CT scans to evaluate the buco maxillofacial region with significantly lower radiation doses than conventional CT scans.

Acknowledgements

We would like to thank GEP / HUSM; Animati PACS; Institute of Radiology São Lucas; Clinic CROM; Tecnoparque of Santa Maria; Stallo 3D solutions.

References

- [1] C.M. Heyer, P.S. Mohr, S.P. Lemburg, S.A. Peters, and V. Nicolas, Image quality and radiation exposure at pulmonary CT angiography with 100- or 120-kVp protocol: prospective randomized study, *Radiology* **245**(2) (2007), 577–583.
- [2] T. Buxi, A. Yadav, K.S. Rawat, and G.S. Ghuman, Effect of iterative reconstructions in low dose computed tomography, *J Biomed Graph Comput* [Internet] **4**(3) (2014), 1–9. Available at: <http://www.sciedu.ca/journal/index.php/jbgc/article/view/4282>
- [3] H. Alsleem and R. Davidson, Factors affecting contrast-detail performance in computed tomography: A review, *J Med Imaging Radiat Sci* [Internet] **44**(2) (2013), 62–70. Available at: <http://dx.doi.org/10.1016/j.jmir.2012.12.001>
- [4] D.G. Garib, R. Raymundo Jr., M.V. Raymundo, D.V. Raymundo, and S.N. Ferreira, Tomografia computadorizada de feixe cônico (Cone beam): entendendo este novo método de diagnóstico por imagem com promissora aplicabilidade na Ortodontia, *Rev Dent Press Ortod e Ortop Facial* [Internet] **12**(2) (2007), 139–156. Available at: <http://rorrj.com.br/atualizacao-cientifica/Tomografia-computadorizada-de-feixe-conico-RORRJ.pdf>
- [5] W.C. Scarfe and A.G. Farman, What is Cone-Beam CT and How Does it Work? *Dent Clin North Am* **52**(4) (2008), 707–730.
- [6] J.P. Dillenseger, J. Matern, C. Gros, F. Bornert, C. Goetz, J.M. Le Minor, and et al, MSCT versus CBCT: evaluation of high-resolution acquisition modes for dento-maxillary and skull-base imaging, *Eur Radiol* [Internet] **25**(2) (2015), 505–515. Available at: <http://link.springer.com/10.1007/s00330-014-3439-8>
- [7] C. Zacharias, A.M. Alessio, R.K. Otto, R.S. Iyer, G.S. Philips, J.O. Swanson, and et al, Strategies to Lower Pediatric CT Radiation Dose, *Am Roentgen Ray Soc* [Internet] **200**(5) (2013), 951–956. Available at: <http://www.ajronline.org/doi/pdf/10.2214/AJR.12.9026>
- [8] J. Dalmazo, E.J. Jorge, M.A.C. Brocchi, P.R. Costa, and P.M. de A. Marques, Otimização da dose em exames de rotina em tomografia computadorizada: estudo de viabilidade em um Hospital Universitário, *Radiol Bras* **2** (2013), 241–248.
- [9] K.J.L. Jung, K.S. Lee, S.Y. Kim, T.S. Kim, Y.S. Pyeun, and J.Y. Lee, Dose, and usefulness for evaluation of bronchiectasis, *Invest Radiol* **35**(9) (2000), 557–563.
- [10] M.C.B. Godoy, S.L. Heller, D.P. Naidich, B. Assadourian, C. Leidecker, B. Schmidt, and et al, Dual-energy MDCT: Comparison of pulmonary artery enhancement on dedicated CT pulmonary angiography, routine and low contrast volume studies, *Eur J Radiol* **79**(2) (2011), 11–7.

Address for correspondence

Gustavo Nogara Dotto, DS, MS, Cd. e-Health Unit, University Hospital of Santa Maria Roraima Avenue, 1000. Santa Maria, RS, Brazil. Zip Code: 97105-900. Phone: +55-55-32131502

Successful System, Incomplete Data: Caveats in Reusing Activity Data from Emergency Department Whiteboards

Morten Hertzum

Royal School of Library and Information Science, University of Copenhagen, Copenhagen, Denmark

Abstract

During the daily treatment of patients large quantities of data are recorded in electronic health records (EHRs). Compared to data in paper records, these EHR data are easily available for reuse in research and quality improvement. However, the opportunities for reuse depend on the quality of the data. In this study we analyze the completeness with which main treatment activities are recorded on emergency department (ED) whiteboards and whether completeness varies with the severity of the patients' condition. Data from 381,231 ED visits show that after the whiteboard had been in successful use for several years the clinicians recorded four of the five main treatment activities with a completeness of less than 50%. Completeness tended to increase with three indicators of the severity of the patient's condition: triage level, length of stay, and patient age. We conclude that the low completeness of the activity data probably prevents most types of reuse.

Keywords:

Quality Improvement; Electronic Health Records; Emergency Medical Services.

Introduction

The substitution of electronic for paper records has been a major development in healthcare organizations over the past decades. Electronic health records (EHRs) contain large quantities of data. These data are recorded during the daily treatment and care of patients but, subsequently, become available for other uses. EHRs have been associated with unprecedented opportunities for improving healthcare through the reuse of EHR data for clinical research, quality improvement, and other data-driven efforts to learn from past events [7]. However, such learning requires quality data. In this regard Weiskopf and Weng [16] contend that EHRs have “led not to improvements in the quality of the data being recorded, but rather to the recording of a greater quantity of bad data” (p. 144). This study focuses on the data recorded on electronic whiteboards in emergency departments (EDs), which are the common entry point to hospital treatment for most patients.

ED whiteboards contain data about the patients and their flow through the ED, such as the patients' time of arrival, triage level, current treatment activity, responsible clinician, and lab test results. Ready access to these data is central to the coordination of ED work and to each ED clinician's sense of overview [8; 11]. In addition to supporting work in the ED as it unfolds, the whiteboard recordings provide opportunities for learning. These learning opportunities include determining whether the recordings can be used for forecasting temporal patterns in patient arrivals, determining waiting and

boarding times, identifying bottlenecks in the patient flow, assessing whether length of stay (LOS) targets are met, comparing the patient mix of EDs, and researching how the coordination of ED work is accomplished. Learning about these issues is important because the ED is a stressful environment for patients [5] as well as clinicians [4]. Improved knowledge of patient flows and resource bottlenecks can help EDs streamline work procedures, dynamically match resources to patient volumes, and prevent ED crowding [13]. We are involved in such learning efforts in the EDs in Region Zealand, one of the five healthcare regions in Denmark [e.g., 9; 11]. In the present study we analyze almost three years of log data from the whiteboards in the four EDs of the region *to assess the completeness with which main treatment activities are recorded and whether completeness varies with indicators of the severity of the patient's condition*. Information about the time spent on the different treatment activities is, for example, crucial to understanding how crowding arises and when countermeasures are required. Thus, the generation of accurate forecasts of crowding presupposes reasonably complete recordings of the treatment activities.

Weiskopf and Weng [16] identify three fundamental dimensions of data quality – completeness, correctness, and currency – and two auxiliary dimensions (concordance and plausibility) that often serve as proxies for the fundamental dimensions when they cannot be assessed directly. Completeness, the focus of this study, concerns whether a piece of data about a patient is present in the EHR. Correctness concerns whether the data that are present in EHRs are also accurate. And currency concerns whether the data are representative of the patient's state at a desired point in time, often interpreted as whether the data are recorded in the EHR within a reasonable period of time following measurement. Previous studies have demonstrated that the data quality is modest for multiple types of EHR data [2; 7; 12; 14]. For example, Chan et al. [2] reviewed the completeness of blood-pressure recordings across multiple studies and found that the number of complete recordings ranged from 0.1% to 51%. In addition, Brennan et al. [1] found that British hospital statistics for 2009-2010 showed nearly 20,000 adults attending pediatric outpatient services and over 8,000 males admitted to gynecology inpatient wards. Several reasons have been proposed to explain modest data quality, including habits, lack of time, failure to capture data that became available to clinicians who were not part of the department, and transcription errors in transferring data from paper charts to EHR [e.g., 12; 14]. At root, data quality suffers from differences in priorities between day-to-day clinical work and work such as research and quality improvement. In day-to-day clinical work data quality is secondary to patient treatment.

Method

The four EDs in Region Zealand were part of medium-sized hospitals and collectively served a population of approximately 817,000 citizens. Prior to conducting the study we obtained approval from the healthcare region.

The Whiteboard

The four EDs installed the same electronic whiteboards, which were introduced in December 2009 (ED1), January 2010 (ED2), January 2011 (ED3), and May 2011 (ED4), respectively. The whiteboard supplemented the electronic patient record by providing procedural information about the patients. Some of the whiteboard data, such as lab test results, were automatically updated when new data became available, but the majority of the whiteboard data were entered and updated manually. In this study we focused exclusively on data that were entered and updated manually. Manual data entry and updating could be done on the wall-mounted whiteboard displays (see Figure 1) as well as on any other computer in the EDs.



Figure 1 – The whiteboard at ED4. For each patient the whiteboard gives one row of information, including time of arrival, triage level, first name, age, responsible physician, current treatment activity, lab-test results, and next stop.

Before turning to analyze the completeness of the whiteboard data and thereby their reuse potential, it is important to note that the whiteboard was successful with respect to its primary purpose of supporting the clinical work in the EDs. A physician at ED3, for example, said in an interview that the whiteboard “gives a great overview. I cannot imagine that we could do without it.” A nurse at ED1 expressed agreement: “We use it all the time”. More formally, a survey at ED1 and ED2 showed that the clinicians experienced an improved overview of their work when the electronic whiteboard replaced the former dry-erase whiteboards [8]. The survey also showed that, to a larger extent, the electronic whiteboard made information available where and when clinicians needed it.

Log Data

All changes of the whiteboard content were automatically logged. For the purpose of this study the whiteboard vendor, Imatis, produced a version of the logs from which all patient names, clinician names, and other information that might identify persons had been removed. These anonymized log data covered the three-year period 2012-2014. However, we had to discard the periods January 2013-January 2014 (ED1) and November 2013-January 2014 (ED2-ED4) from the analysis because they contained long intervals of no data.

After also removing 741 outliers (defined as ED visits longer than seven days, i.e. more than 50 times the median length of stay), the dataset comprised 381,231 ED visits. Table 1 shows the division of the visits into EDs and years.

Table 1 – The 381,231 ED visits divided onto ED and year

ED	2012 (Jan-Dec)	2013 (Jan-Oct)	2014 (Feb-Dec)
ED1	33,040	-	30,719
ED2	40,445	32,844	37,396
ED3	32,677	28,527	34,304
ED4	38,670	32,628	39,981

The data for the 381,231 ED visits consisted of over 10 million log entries, each documenting an event that changed the whiteboard content. A log entry contained a timestamp, an event type, any values associated with the event, and a system-generated identifier of the visit to which the event pertained. For example, the event type ‘WAITING_FORChanged’ along with the event value ‘Nurse’ indicated that the patient was now waiting to be seen by a nurse. This event type marked the start of a treatment activity in the patient’s progress through the ED workflow. Across the EDs different sets of treatment activities (i.e., different sets of event values) were used for indicating the stages of this workflow. However, the workflows of the EDs shared five main activities: triage, waiting to be seen by a nurse, waiting to be seen by a physician, examination (by a junior physician), and review (by a senior physician). No other treatment activity was recorded more consistently than these five main activities. The log data also contained information about changes in, among other things, responsible physician and lab test results; but in this study we focus on the five main treatment activities because the flow of the patients through the ED is important to, for example, forecasting and preventing crowding.

Results

The ability to record the patients’ progress through the ED workflow was included on the whiteboard because at-a-glance access to this information was deemed important to the ED clinicians’ overview of their collective work. Yet, even the five main treatment activities were recorded for only a subset of the ED visits (see Table 2). The main activity recorded most and least often differed across the EDs. In ED3 and ED4 waiting to be seen by a nurse was recorded for 74% and 76%, respectively, of the patients in 2014 and for similarly high percentages of patients in the preceding years. The only other activity recorded for the majority of the patients was waiting to be seen by a physician in ED2 (52% in 2014). Conversely, the activity of triage was recorded for less than 1% of the patients in ED1 and ED3. For all EDs at least one of the five activities was recorded for no more than 20% of the patients.

It could be hoped that the completeness of the recordings increased over time because the clinicians appreciated the improved overview or became more conscientious in their use of the whiteboard. The data provided little ground for such hopes. Rather, the trends in the data from 2012 to 2014 showed a mixed picture with nine increasing trends, four decreasing trends, and seven unchanging trends (see Table 2). It should also be noted that prior to 2012 the whiteboard had been in operational use at the EDs for between half a year and two years; thus, work procedures involving the whiteboard had had time to stabilize.

Table 2 – Completeness of the recordings of the five main treatment activities

Treatment activity	2012		2013		2014		Trend ^a
	N	%	N	%	N	%	
ED1							
Triage	12	0.04	-	-	3	0.01	→
Nurse	267	0.81	-	-	143	0.47	→
Physician	5,552	17	-	-	6,404	21	↗
Examination	11,292	34	-	-	11,252	37	↗
Review	8,264	25	-	-	8,935	29	↗
ED2							
Triage	10,028	25	9,521	29	11,871	32	↗
Nurse	4,604	11	3,565	11	3,435	9	↘
Physician	19,616	49	14,919	45	19,459	52	↗
Examination	14,217	35	11,161	34	11,217	30	↘
Review	17,772	44	14,436	44	17,228	46	↗
ED3							
Triage	37	0.11	37	0.13	22	0.06	→
Nurse	24,270	74	17,896	63	25,436	74	→
Physician	10,443	32	9,576	34	12,024	35	↗
Examination	9,224	28	7,732	27	10,269	30	→
Review	1	0.00	1	0.00	0	0.00	→
ED4							
Triage	9,867	26	5,815	18	7,398	19	↘
Nurse	27,366	71	25,018	77	30,201	76	↗
Physician	2,770	7	10,332	32	15,569	39	↗
Examination	8,531	22	7,163	22	9,577	24	→
Review	7,793	20	5,112	16	6,091	15	↘

^a Trend in completeness from 2012 to 2014: ↗ - an increase of more than 2.00 percentage points, ↘ - a decrease of more than 2.00 percentage points, → - a change of at most 2.00 percentage points.

The first activity in the ED workflow, triage, illustrated the important point that failing to record triage as the current treatment activity for a patient did not indicate that the patient was not triaged. The whiteboard gave the triage code for 77% (ED1), 37% (ED2), 10% (ED3), and 40% (ED4) of the patients. For all four EDs the number of patients with a triage code exceeded the number of patients for which triage was recorded as the current treatment activity. Probably, triaging a patient and recording the triage code were experienced as more clinically relevant by the ED clinicians than recording that their current treatment activity was to triage the patient, especially because triage was a brief procedure.

An important consideration in assessing the possibilities for data reuse is whether completeness varies with indicators of the patient's condition. Tables 3 to 5 show how completeness varied with three indicators of the severity of the patient's condition. To save space each table gives the data for only one of the four EDs.

Table 3 shows how completeness varied with triage level in ED1. Waiting to be seen by a physician was more often recorded as the current treatment activity for patients triaged 4 and 5 (i.e., the most severe cases) and examination and review were most often recorded for patients triaged 2 and 3. It might have been more clinically relevant to record the activity in progress for the patients triaged 2 and 3 because they were in the ED longer than the other patients. For example, many of the patients triaged 4 and 5 were quickly transferred to inpatient departments for specialist treatment. In ED1 the completeness of the treatment activities triage and waiting to be seen by a nurse was largely unaffected by triage level. The patterns for physician and examination were roughly similar in the other EDs, whereas the patterns for triage, nurse, and review were different. For example, waiting to be seen by a nurse in ED4 was recorded for 80% of the patients triaged 2 and 3 but only for about half as many at the other triage levels.

When completeness varied with the triage level, it tended to be by higher completeness at medium or high triage levels.

Length of stay (LOS) directly indicates a patient's demand on ED resources and indirectly indicates the severity of the patient's condition. Table 4 shows that the activity of triage tended to be recorded more often for the patients that ended up staying longer in ED2. Waiting to be seen by a physician was most often recorded for the patients who stayed 3-8 hours; ED3 was similar in this respect. In ED2, the activity of review was recorded as the current treatment activity for fewer and fewer patients as LOS increased. The completeness of the review recordings also varied systematically with LOS in ED1 and ED4 but in the opposite direction: completeness increased with increasing LOS. Waiting to be seen by a nurse was the only treatment activity the recording of which did not vary appreciably with LOS in any of the EDs. The four other treatment activities were, in one or two of the EDs, recorded more often with increasing LOS. The only instance of a decreasing trend was for review in ED2.

Table 3 – Completeness of treatment-activity recordings for ED1, by triage level

Triage level	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
1	1482	0.00	0.00	7	34	38
2	11857	0.01	0.03	12	72	50
3	12131	0.02	0.19	15	70	56
4	23230	0.05	1	28	16	12
5	397	0.00	5	24	1	2

Note. The table includes only the 49,097 (77%) ED1 visits for which the triage level was available.

Because older patients tend to be weaker than younger patients, we use age as a third indicator of the severity of the patient's condition. In ED4, waiting to be seen by a nurse was recorded less often with increasing patient age, while waiting to be seen by a physician, examination, and review were recorded more often with increasing patient age (see Table 5). In the other EDs waiting to be seen by a nurse was recorded

about equally often for the different age groups but the pattern for waiting to be seen by a physician and for examination resembled that in ED4. In ED1 the pattern for the activity of review also resembled that in ED4, but in ED2 it was reversed (as it was for LOS). The pattern that completeness tended to increase with increasing patient age was stronger than the patterns for triage and LOS.

Table 4 – Completeness of treatment-activity recordings for ED2, by length of stay (LOS)

LOS (hours)	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
0-2	63,490	22	9	41	27	54
3-5	32,087	35	12	61	41	39
6-8	5,533	38	14	67	47	27
9-11	1,896	32	9	55	39	19
12-	7,679	42	11	53	37	11

Table 5 – Completeness of treatment-activity recordings for ED4, by patient age

Patient age (years)	ED visits	Triage %	Nurse %	Phys. %	Exam. %	Review %
0-19	12,982	16	82	15	8	7
20-39	15,753	28	77	30	24	19
40-59	19,265	29	73	37	31	23
60-79	21,026	27	71	42	39	29
80-	11,531	17	68	43	40	30

Note. The table includes only the 80557 (72%) ED4 visits for which the patient age was available.

For all three indicators the overall, but not unanimous, pattern was that completeness increased when the indicator pointed toward patients with more severe conditions.

Discussion

After the whiteboard had been in successful use in the EDs for 2-4 years the five main treatment activities were recorded in 2014 for an average of 13% (triage), 40% (nurse), 37% (physician), 30% (examination), and 23% (review) of the patients. These averages hide considerable variation across the EDs; but no ED recorded more than one treatment activity with a completeness of more than 50%, and all EDs recorded at least one treatment activity with a completeness of less than 20%. The substantial incompleteness in the recording of the patients' current treatment activity is the result of a constant tension between treating patients and documenting treatments. Another result of this tension is that the incompleteness is not randomly distributed. Rather, completeness tended to increase with increasing triage level, LOS, and – most strongly – patient age. We make two conclusions from the analysis:

- The incompleteness of the activity data is substantial, probably preventing most types of reuse.
- Reusing the data incurs a bias toward patients with more severe conditions.

The former finding confirms previous research [2; 14], thereby indicating that EHRs should mainly be assessed on the basis of how well they fulfill their primary purpose of supporting day-to-day clinical work. The latter finding acknowledges that the mere existence of large quantities of real-world EHR data provides impetus for their reuse [7]. If

the data are reused it is important to be aware of their limitations. The bias toward patients whose triage level, LOS, and age indicate a more severe condition extends the finding by Rusanov et al. [15] that EHRs contain more data about the sicker patients.

Part of the reason for data incompleteness is that the decision about whether to record the data is often left to the clinicians' discretion. This practice acknowledges the primacy of their day- to-day clinical work. Somewhat surprisingly it remains unclear in many studies of the quality of EHR data [e.g., 2; 12] whether it was mandatory for the clinicians to record the data. Frequently, transitional artifacts fill a gap between the clinical work and the formal documentation of it [3]. Such transitional artifacts hold procedural information important in performing the work, but at the same time the transitional artifacts are exempted from the formal documentation requirements. The whiteboard in the present study is an example of a transitional artifact. That is, the clinicians were not formally required to keep the whiteboard current. While this may contribute to explaining the incomplete data, a formal requirement to record the treatment activities will not necessarily result in complete data. For example, Granlien and Hertzum [6] found that none of eight mandated work procedures associated with an electronic medication record were followed consistently by more than 48% of the wards at the studied hospitals.

Another reason for the incomplete data is that the treatment activities had to be recorded manually. Activity data are, for example, pertinent in moving the modelling of ED crowding beyond models based solely on when patients arrive in the ED [10]. Thus, the introduction of a tool that forecasts ED crowding on the basis of activity data might motivate the clinicians to record these data more consistently. Alternatively, it might be possible to derive activity data automatically from other data. For example, the recording of a patient's triage level indicates the end of the activity of triage and is, in most cases, also a good indicator that the patient has now transitioned to the activity of waiting to be seen by a nurse. The triage level was not recorded for all patients, but in all EDs it was recorded more often than that the patient was waiting to be seen by a nurse. Thus, the requirement for manual data entry can probably be reduced by deriving additional activity data from other whiteboard data or from EHR data. While automatic data derivation will likely improve completeness [2], data correctness may suffer because manual data entry likely captures some nonsensical data. In balancing manual data entry against automatic data derivation it should also be considered that automatic data derivation frees clinician time for other activities.

Finally, the bias of the recorded data toward patients with more severe conditions is reassuring from a clinical point of view because it suggests that the clinicians attend more to the patients who are more in need of clinical attention. It may, however, be impossible to compensate for this bias when the data are reused, thereby confounding any analyses performed by reusing the data.

Limitations

Two limitations should be remembered in interpreting the results of this study. First, the data are from EDs in one healthcare region of one country. While the four EDs show that the results of the study are not peculiar to one ED, it would be valuable to replicate the study in other countries with other healthcare systems. Second, we can merely speculate about the reasons for the incomplete recording of the

treatment activities. While the log data quantify the magnitude of this incompleteness, interviews or other qualitative data would be needed to explain why the clinicians like the whiteboard but often leave the recording of the treatment activities incomplete.

Conclusion

Changes to the content of electronic ED whiteboards are logged and thereby available for later inspection and reuse. Such log data provide opportunities for forecasting ED crowding, identifying bottlenecks in the ED workflow, and – more generally – for reusing EHR data for the purposes of research and quality improvement. In this study we have analyzed the completeness with which five main treatment activities are recorded on the ED whiteboards in a Danish healthcare region. We find that the low completeness of the activity data probably prevents most types of reuse and that, if reused, the activity data incur a bias toward the patients with the more severe conditions as indicated by triage level, length of stay, and patient age. The incomplete activity data cannot be explained by the whiteboard being disliked by the ED clinicians or unused for its primary purpose of supporting day-to-day ED work. On the contrary, the whiteboard had been in successful use in the EDs for several years. It is in spite of successful primary use that the incompleteness of the activity data probably prevents most secondary uses. We point to non-mandated use and manual data entry as reasons that contribute to the poor data quality.

Acknowledgements

This study is part of the Clinical Communication project, which is a research and development collaboration between Roskilde University, University of Copenhagen, Imatis, and Region Zealand. The two interview quotes in the Method section are from interviews conducted in collaboration with Jesper Simonsen and Arnvør á Torkilsheyggi, both from Roskilde University. Special thanks are due to Rasmus Rasmussen at Imatis for making the anonymized version of the log data.

Conflicts of interest: None to declare.

References

- [1] L. Brennan, M. Watson, R. Klaber, and T. Charles, The importance of knowing context of hospital episode statistics when reconfiguring the NHS, *British Medical Journal* **344** (2012), e2432.
- [2] K.S. Chan, J.B. Fowles, and J.P. Weiner, Electronic health records and the reliability and validity of quality measures: A review of the literature, *Medical Care Research and Review* **67** (2010), 503-527.
- [3] Y. Chen, Documenting transitional information in EMR, in: *Proceedings of the CHI2010 Conference on Human Factors in Computing Systems*, ACM Press, New York, 2010, pp. 1787-1796.
- [4] L. Flowerdew, R. Brown, S. Russ, C. Vincent, and M. Woloshynowych, Teams under pressure in the emergency
- [5] J. Gordon, L.A. Sheppard, and S. Anaf, The patient experience in the emergency department: A systematic synthesis of qualitative research, *International Emergency Nursing* **18** (2010), 80-88.
- [6] M.S. Granlien and M. Hertzum, Barriers to the adoption and use of an electronic medication record, *Electronic Journal of Information Systems Evaluation* **15** (2012), 216-227.
- [7] W.R. Hersh, M.G. Weiner, P.J. Embi, J.R. Logan, P.R.O. Payne, E.V. Bernstam, H.P. Lehmann, G. Hripcsak, T.H.artzog, J.J. Cimino, and J.H. Saltz, Caveats for the use of operational electronic health record data in comparative effectiveness research, *Medical Care* **51** (2013), S30-S37.
- [8] M. Hertzum, Electronic emergency-department whiteboards: A study of clinicians' expectations and experiences, *International Journal of Medical Informatics* **80** (2011), 618-630.
- [9] M. Hertzum, Patterns in emergency-department arrivals and length of stay: Input for visualizations of crowding, *Ergonomics Open Journal* **9** (2016), 1-14.
- [10] M. Hertzum, Forecasting hourly patient visits in the emergency department to counteract crowding, *Ergonomics Open Journal* **10** (2017), to appear.
- [11] M. Hertzum and J. Simonsen, Visual overview, oral detail: The use of an emergency-department whiteboard, *International Journal of Human-Computer Studies* **82** (2015), 21-30.
- [12] W.R. Hogan and M.M. Wagner, Accuracy of data in computer-based patient records, *Journal of the American Medical Informatics Association* **4** (1997), 342-355.
- [13] N.R. Hoot and D. Aronsky, Systematic review of emergency department crowding: Causes, effects, and solutions, *Annals of Emergency Medicine* **52** (2008), 126-136.
- [14] A. Parsons, C. McCullough, J. Wang, and S. Shih, Validity of electronic health record-derived quality measurement for performance monitoring, *Journal of the American Medical Informatics Association* **19** (2012), 604-609.
- [15] A. Rusanov, N.G. Weiskopf, S. Wang, and C. Weng, Hidden in plain sight: Bias toward sick patients when sampling patients with sufficient electronic health data for research, *BMC Medical Informatics and Decision Making* **14** (2014), 51.
- [16] N.G. Weiskopf and C. Weng, Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research, *Journal of the American Medical Informatics Association* **230** (2013), 144-151.

Address for correspondence

Morten Hertzum, University of Copenhagen, Njalsgade 76, Bldg. 4, Copenhagen, Denmark, hertzum@hum.ku.dk.

Improving Patient Safety Through the Design and Development of a Computerized Provider Order Entry for Parenteral Nutrition Linked to a Barcode Medication Administration Record

Oscar I Jáuregui, Lucila Bruchanski, Daniel A Rizzato Ledo, Carlos M Otero, Daniel R Luna

Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Among adverse events related to medication errors, the defects in parenteral nutrition administration pose a special threat to patient safety. Two high impact strategies to reduce these errors require implementing a Computerized Provider Order Entry and the use of bedside bar-code verification prior to medication administration. In this study, we share the deep field analysis of the current workflow performed to include inpatient bedside barcoding verification for parenteral nutrition administration in a large academic health center. Then, we propose a process optimization and a new parenteral nutrition ordering tool embedded in the prescription module. Structuring physicians' ordering and the bar-code verification of administration can increase patient safety. The next steps would involve the creation of a Clinical Decision Support System to improve patient nutrient goals.

Keywords:

Electronic Health Records; Parenteral Nutrition; Patient Safety.

Introduction

Medication errors have been reported as the most common causes for alerts in critical incident reporting systems [1]. Also adverse events related to inpatient medication errors are a serious menace to patient safety [2] having an incidence of 6,5%. One third of them take place during prescription and another third occur during medication administration [3]. It was found that the parenteral nutrition (PN) administration was defective in 3,2% of the cases [4].

There is growing evidence that Health Information Technology (HIT) for medication safety — classified as a socio-technical system, due to the complexity of the medication processes and the required coordination of various professional groups [5] — can help to overcome this problem. Computer Provider Order Entry (CPOE) implementation was associated with a reduction in serious medication errors by 55% [6]. In addition, the use of bedside bar coded medication administration technology (BCMA) diminished errors in medication administration from 41% to 27%, and proved to be cost-effective [7]. Although the feeding ordering functionalities of commercial Electronic Health Records (EHR) are heterogeneous [8], they can help physicians by providing clinical decision support systems (CDSS), therefore decreasing the risk of drug interactions and transcription errors, among others [9].

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit health care academic center founded in 1853, with over 2,700 physicians, 2,700 other health team members (including 1,200

nurses) and 1,800 administrative and support employees. The HIBA has a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 homecare beds, 25 outpatient clinics and 150 associated private practices located in Buenos Aires city and its suburban area. It has a Health Maintenance Organization (Plan de Salud) that covers more than 150,000 people and also provides health services to another 1,500,000 people who are covered by affiliated insurers. Between 2013 and 2014, over 45,000 inpatients were admitted to its hospitals, there were 45,000 surgical procedures (50% ambulatory) and 3,000,000 outpatient visits.

Since 1998, the HIBA runs an in-house developed health information system, which includes clinical and administrative data [10]. Its Electronic Health Record (EHR) system called Italica, is an integrated, modular, problem oriented and patient centered system that works in different clinical settings (outpatient, inpatient, emergency and home care. It has been recently certified by the HIMSS as level 6+ in the Electronic Medical Record Adoption Model (EMRAM), being the first hospital in Argentina and the second in Latin America reaching this stage [11] [12].

The HIBA achieved Joint Commission International safety and quality certification in 2015. One major addressed issue was the correct identification of patients. To this end, in 2010 the Department of Health Informatics developed and implemented an inpatient bracelet identification system with barcoding technology. This code can be scanned to verify patient identity before every intervention (patient transfer, medication administration or other procedures).

Nowadays, just Adult Intensive Care Unit fulfills the process of scanning the tag of every medication before administration. As suggested by the Institute of Medicine, the Hospital Italiano de Buenos Aires is using HIT to enhance patient safety [13]. As CPOE and CDSS were fully implemented for most medications, and in order to certificate EMRAM level 7, we decided to take a step forward and apply bedside BCMA in every inpatient setting, for each substance administration, including parenteral nutrition.

An average of 80 PN bags per day is prescribed to inpatients in our institution. Half of them are for adult patients and the other half are distributed evenly between pediatric and neonatal patients. The bags are provided daily by an outsourced vendor, who produces them with an automated compounding device (ACD) that needs the manual entry of the prescription data.

Parenteral nutrition is a complex type of medication that has a specific process and requires special attention to succeed in closed loop administration. We analyzed and modeled current processes in order to define IT system requirements [14], and then designed and developed a new software tool according to

health IT medication safety guidelines [5] in order to prevent from technology-induced errors [15]. These actions take part of a broad overall project for BCMA implementation, included in the portfolio for Electronic Medical Record Adoption Model Level 7 certification.

The two objectives of this paper are to describe the process modeling and optimization and to show the participative and user centered design of a structured PN ordering module linked to BCMA.

Methods

Process modeling and optimization

We sought to analyze and model the processes and workflows involved in the prescription, preparation, validation and administration of the PN bags in our institution. We collected the data based on a three stage methodology, according to the work published by Granja et al. [14], as follows:

Stage 1: We performed database queries to collect quantitative and qualitative historical data of the number of NPT prescriptions, prescribing medical services, items and structure of prescription, and the differences between services in the last 6 months.

Stage 2: We subsequently carried out field observations in the different areas involved in the process (wards, medical offices, central pharmacy, satellite pharmacies and nurse offices), with unstructured interviews with the involved health care workers.

Stage 3: We finally accomplished in-depth interviews with key stakeholders (physicians, pharmacists, nurses, and the outsourced vendor).

Afterwards the processes were plotted with flowcharts and task breakdown, and worked under project management methodology [16].

We then proposed certain modifications to the current process in order to achieve BCMA requirements. These changes were agreed upon with the end users to avoid change resistance in the subsequent implementation.

Prescription interface design

We intended to create a new structured parenteral nutrition ordering functionality within our CPOE. According to current state of art in safe IT design and medication safety [5,15] we performed an iterative, participative and user-centered design, and followed accepted design guidelines, in order to minimize further technology induced errors. Some adjustments were also included in the medication list visualization for physicians and nurses, in order to adequately represent the PN data being entered in the novel module.

Results

Process analysis

We analyzed a 6 month period (March 2016 to August 2016) of free text TPN prescriptions within our CPOE. A total of 15101 prescriptions were made (mean=83 per day) by 4 different medical services. We qualitatively analyzed the variables involved in the prescriptions and found that type of TPN, method of infusion, route, infusion rate, start time, cycling periods and composition were the main items. We also found profound differences between services in the prescriptions' text, namely that one service only specified route and infusion rate, while another one explicitly wrote bags' composition and every cycling period.

We carried out observations in the different areas where the process developed, and had interviews with the corresponding personnel to better understand the steps being observed. Then we performed in-depth interviews with key stakeholders from the different areas involved in order to better understand the process, and dig in the aspects that were found to be inefficient, that could lead to a breach in patient safety, or that would not allow us to implement BCMA.

The process begins with the prescription of the parenteral nutrition bags (Figure 1). It is ordered by the physicians after the daily rounding. The prescription is made twice: first they write the composition and the posology in free text in the EHR. Later, a structured paper form is filled manually with the patient's identity and the PN desired composition, which is scanned and sent by email or fax to the vendor. The vendor technician inputs the data manually into the compounding machine's system. He performs a validation process to check the components against standardized values. If the formula is correct, the bag is prepared and then sent to our institution, labeled with the patient's information and a production barcode identifier. The PN bag is received at the central pharmacy where a pharmacist searches for the patient in our EHR, and validates manually the right composition, comparing it with the physician's prescription. Afterwards, the bag is dispensed to each inpatient unit and stored in a special refrigerator in the satellite pharmacy. The nurse picks up the bag (checking the patient's identity in the bag label), connects it to the patient and starts the administration through an infusion pump. Finally, she documents manually her actions in our electronic medication administration record (eMAR).

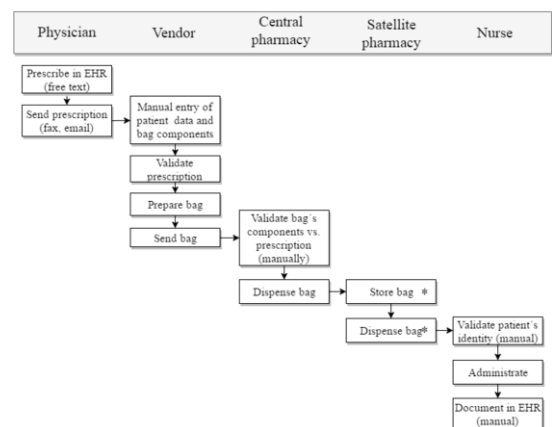


Figure 1 - Process analysis. *: steps only done in intensive care units. In open units the bag is dispensed directly to the nurse.

Process optimization

We proposed some process modifications to achieve closed loop administration (Figure 2), namely:

1. Parenteral nutrition posology has to be structured to standardize the communication between physicians and nurses, and to allow right route, time and dose (volume) automatic verification.
2. As the outsourced vendor requires a handwritten PN prescription sheet, we have to include univocal patient identity information together with a barcode that contains the patient's hospital identification number.
3. The patient's hospital identification barcode should be scanned by the vendor technician, and their system would verify the correctness through a web service.

4. The nurse should scan the patient's wristband and then the bag's production barcode identifier prior to administration. If the bag and prescription matches the patient, the system will allow her to administrate the bag, otherwise an alert would appear.
5. The nurse records the administration synchronously.

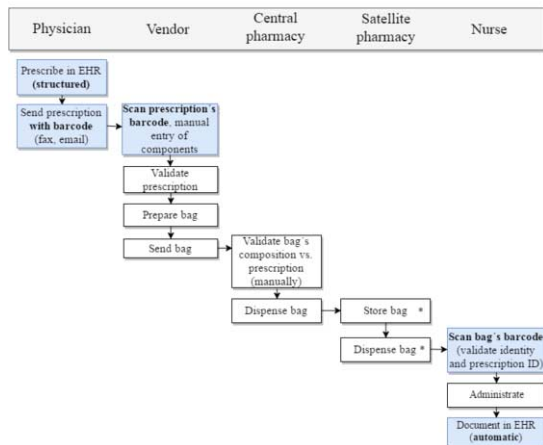


Figure 2 - Process optimization. (When a step has a modification, box is painted blue). *: steps only done in intensive care units. In open units the bag is dispensed directly to the nurse.

Prescription module design

We designed a new PN ordering module for our EHR that should be accessible for physicians from the CPOE.

The design started with the identification of the main tasks of the prescription process, made by direct observation and interviews. Then, a mock-up of the user interface (UI) was prototyped (Figure 3) and assessed for heuristic evaluation by

a usability expert, then presented to key users, who qualitatively tested it in order to undermine major usability flaws. After we analyzed user's feedback, a number of improvements in the UI and new functionalities were presented to users for acceptance. After iteration, a final prototype was exhibited to users, who were able to complete tasks without any usability problems, and with good overall acceptance (Figure 4).

The interface works as follows: the professional initially selects the type of preparation (compounded or ready to use), the route and the administration method.

When the provider chooses a ready to use preparation, the system displays the products available in the pharmacy. We decided to show the brand name, as the generic components were confusing for users during the test sessions. The infusion method can be selected between cycling or constant infusion rate with a radio-button. If constant infusion rate is selected, provider has to enter the volume to infuse, the infusion rate and the time to start administration. If the users choose the cycling method the system displays a schedule with start and stop time, and corresponding infusion rate has to be entered. The partial and total volumes are automatically calculated (functionality extracted from testing session's analysis, as providers had difficulties calculating these volumes). When the provider chooses a compounded preparation, he must describe its formulation in a free text field. In both cases, an optional text field is offered for additional observations.

From this interface, the user can print an empty PN form filled with the patient's name, date of birth, date of prescription, hospital ID number and barcode. The physician has to add manually the PN formulation and sign it before scanning and sending it to the vendor.

The PN orders were highlighted in the patient medication list, to avoid misunderstandings between nurses and physicians.

Figure 3 shows two mockups of the TPN prescription interface. The left mockup shows the initial state with all elements displayed at start, creating confusion. The right mockup shows the interface after user input, where the 'Ready to use' type is selected, the product is 'Nutriflex', the method is 'Constant IR', and the time is set to 12:17. The left mockup has a 'Type' dropdown with 'Compounded' selected, a 'Product' dropdown with 'Compounded' selected, a 'Route' dropdown with 'central' selected, a 'Volume' dropdown with 'Select an option' selected, and a 'Method' dropdown with 'cycilate' selected. The right mockup has a 'Type' dropdown with 'Ready to use' selected, a 'Product' dropdown with 'Nutriflex' selected, a 'Route' dropdown with 'central' selected, a 'Volume' dropdown with 'Select an option' selected, and a 'Method' dropdown with 'Constant IR' selected. Both mockups have a 'Start' and 'Stop' time field, a 'rate' field, and an 'Observations' text area. The left mockup has a 'Cancel' button, and the right mockup has a 'Cancel' button.

Figure 3 - First mock up of the TPN prescription interface. All elements were displayed at start, creating confusion. Rows in the cycling periods were found difficult to interact with because of the labeling repetition. No composition field was included.

Figure 4 - Final prototype of the TPN prescription module. Elements are displayed sequentially, following the prescription's cognitive workflow. Cycling rows were redesigned to minimize repeated labeling, volumes are auto calculated, and an optional "Composition" field was added.

Discussion

We reviewed the medication cycle process for parenteral nutrition in our institution, which involves an external stakeholder. We encountered some error-prone steps like the free text format ordering in the EHR and the patient's identity verification between both organizations. Moreover, some data was missing in the process, and hindered bedside barcoding implementation. We then proposed some variations to that workflow for increasing patient safety, and closing the data loop. The main improvements in the process were the prescription of parenteral nutrition in a standardized and structured way, for which a specific module was designed to fulfill that requirement; the standardization of patient identity verification with the external supplier; and the scanning of the of the bag's barcode by the nurse prior to administration to validate the 5 rights. Therefore, we expect to diminish significantly the errors due to wrong patient, route or dose (volume), at the prescription and administration phases. Transcription errors can also disappear as the nurse will see the prescription information directly in the eMAR. Nevertheless, as stated in Seidling et al., when implementing a new tool with a process redesign, it is not always possible to separate the benefits of the health information technology intervention from those inherent to the process optimization itself [7].

We designed the system following a user centered and participatory design, meeting user and systems needs. By that, we expect to avoid workflow disruption, and minimize the risk of technology induced errors. Having said that, we should still monitor the human computer interactions and the patient outcomes to prevent workarounds and iatrogenia that might emerge following implementation [17].

A substantial limitation to the process optimization and the implementation of the BCMA is that the outsourced vendor has a non-standard compliant system that hinders the integration with ours. The main barriers for achieving interoperability were that their patient's identification system lacked the security measures needed for clinical purposes, and that the variables they used in the prescription module weren't built following the HL7 standards that run in our system.

Since the software changes necessary for full interoperability will demand further resources, they'll be tackled in a second stage. Thus, the bag's formulation will remain in a free text field until its incorporation to the electronic tool. Additionally, paper is still necessary in our process, and its transcription to the external supplier's system occurs. On the other hand, as health IT for medication safety involves numerous care processes, it should be carefully planned and integrated step by step [5].

Another limitation found was the high heterogeneity of the work patterns among providers, which could influence the IT system's usefulness. These different processes are seen in most of large health institutions, and they are an obstacle for implementing a single IT solution, since it's not recommended to have different work patterns at the same time.

Nowadays we are at the final stage of the development of the prescription module and the communication with the vendor, and implementation is planned for the following months. After implementation we will evaluate its impact in patient safety through specific indicators.

Our software developing group is also currently working in a fully integrated interface between our EHR and the vendor's own system, which allows individual PN component prescription in the CPOE. The PN prescription will be validated at different stages, and will be prepared by the automated compounding machine, preventing data transcription. Moreover, structured PN bag's components allows CDSS integration, thus improving prescriber performance [6] [18] and cost effectiveness [19].

A number of IT solutions have been described in the field of IT solution, specifically BCMA, for medication safety [20][21] but we believe this is the first description of a work of this nature in the South American region.

Conclusions

We analyzed and modeled the medication workflow of parenteral nutrition, and proposed process changes to achieve bedside barcoding readiness to improve medication safety. We

shared the iterative and participatory design of the interface of a novel prescription module for parenteral nutrition orders.

Obstacles were found in integrating an actor outside the organization, without governance over it, into the process, and in creating a single optimized process for the different medical and administrative workflows that existed in the institution.

The software is currently being developed, and future publications will address its implementation, safety outcomes and improvements.

Acknowledgements

We thank the Neonatology, Pediatrics and Nutritional Support services for their time and support of the project. We also thank Juan Ortiz for his collaboration in the mockups design.

References

- [1] Brinkrolf P, Prien T, van A. H., Medication errors – A systematic analysis of CIRS-AINS reports, *Anaesth Intensivmed* 54 (2013) 126–132.
- [2] Poon, E. G., Keohane, C. A., Yoon, C. S., Ditmore, M., Bane, A., Levitzion-Korach, O., et al., Effect of Bar-Code Technology on the Safety of Medication Administration, *N Engl J Med* 362 (2010) 1698–707.
- [3] Bates, D. W., Incidence of Adverse Drug Events and Potential Adverse Drug Events, *JAMA* 274 (1995) 29.
- [4] Li, Q., Kirkendall, E. S., Hall, E. S., Ni, Y., Lingren, T., Kaiser, M., et al., Automated detection of medication administration errors in neonatal intensive care., *J. Biomed. Inform.* 57 (2015) 124–33.
- [5] Ammenwerth, E., Aly, A. F., Brkle, T., Christ, P., Dormann, H., Friesdorf, W., et al., Memorandum on the use of information technology to improve medication safety, *Methods Inf. Med.* 53 (2014) 336–343.
- [6] Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., et al., Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors, *JAMA* 280 (1998) 1311.
- [7] Seidling, H. M., and Bates, D. W., Evaluating the Impact of Health IT on Medication Safety., *Stud. Health Technol. Inform.* 222 (2016) 195–205.
- [8] Vanek, V. W., Ayers, P., Charney, P., Kraft, M., Mitchell, R., Plogsted, S., et al., Follow-Up Survey on Functionality of Nutrition Documentation and Ordering Nutrition Therapy in Currently Available Electronic Health Record Systems., *Nutr. Clin. Pract. Off. Publ. Am. Soc. Parenter. Enter. Nutr.* 31 (2016) 401–415.
- [9] MacKay, M., Anderson, C., Boehme, S., Cash, J., and Zobell, J., Frequency and Severity of Parenteral Nutrition Medication Errors at a Large Childrens Hospital After Implementation of Electronic Ordering and Compounding, *Nutr. Clin. Pract.* 31 (2016) 195–206.
- [10] Luna, D., Plazzotta, F., Otero, C., González Bernaldo de Quirós, F., Baum, A., and Benítez, S., Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires, (2012).
- [11] Noticias Departamento de Informática en Salud del Hospital Italiano, Único Hospital en Argentina en acreditar calidad de Historia Clínica Electrónica, (2016).
- [12] First hospital in Argentina awarded for digital maturity, (n.d.).
- [13] Institute of Medicine (IOM), *Health IT and Patient Safety: Building Safer Systems for Better Care*, (The National Academies Press (US), Washington (DC), 2012).
- [14] Granja, C., Dyb, K., Larsen, E., Roald, S., and Hartvigsen, G., Methodology for Health Care Process Modelling: Bringing the Health Care Complexity into Health IT System Development, *Scand. Conf. Heal. Informatics* (2014).
- [15] Borycki, E., Dexheimer, J. W., Hullin Lucay Cossio, C., Gong, Y., Jensen, S., Kaipio, J., et al., Methods for Addressing Technology-induced Errors: The Current State, *Yearb. Med. Inform.* (2016) 30–40.
- [16] PMI | Project Management Institute, (n.d.).
- [17] Koppel, R., Metlay, J. P., Cohen, A., Abaluck, B., Localio, A. R., Kimmel, S. E., et al., Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors, *JAMA* 293 (2005) 1197.
- [18] Franco, K. A., and O'Mara, K., Impact of Computerized Provider Order Entry on Total Parenteral Nutrition in the Neonatal Intensive Care Unit., *J. Pediatr. Pharmacol. Ther.* 21 (2016) 339–345.
- [19] Zimlichman, E., Keohane, C., Franz, C., Everett, W. L., Seger, D. L., Yoon, C., et al., Return on investment for vendor computerized physician order entry in four community hospitals: the importance of decision support., *Jt. Comm. J. Qual. Patient Saf.* 39 (2013) 312–8.
- [20] Young, J., Slebodnik, M., and Sands, L., Bar Code Technology and Medication Administration Error, *J. Patient Saf.* 6 (2010) 115–120.
- [21] Marini, S. D., and Hasman, A., Impact of BCMA on medication errors and patient safety: a summary., *Stud. Health Technol. Inform.* 146 (2009) 439–44.

Address for correspondence

Oscar Ignacio Jauregui, Hospital Italiano de Buenos Aires. Juan Domingo Perón 4190, C1181ACH. Ciudad de Buenos Aires. Argentina. oscar.jauregui@hospitalitaliano.org.ar

Use of EHR Data to Identify Factors Affecting the Time to Fall

Hyesil Jung^a, Hyeoun-Ae Park^{a,b}

^aCollege of Nursing, Seoul National University, Seoul, South Korea,

^bResearch Institute of Nursing Science, Seoul National University, Seoul, South Korea

Abstract

Although there are many studies of falls occurring in a hospital setting, research on factors affecting time to fall after admission is scarce. It is important for nurses to identify factors contributing to an early fall so that they can pay particular attention to patients with such factors. In this study, patients who sustained a fall were extracted from an adverse event reporting system and narrative nursing records of those hospitalized between January 2015 and May 2016. We used the electronic health records of ten different data sources to extract fall-related variables; the data were integrated according to normalization criteria. Univariate and multiple linear regression analyses were used to identify factors influencing the time to fall from admission. About 49% of fallers fell within the first week after admission. A walking disorder, comorbid disease, intravenous therapy, and arterial lines were related to early falls.

Keywords:

Accidental Falls; Electronic Health Records; Information Storage and Retrieval

Introduction

Falls are one of the most common patient safety problems in hospitals. Falls can lead to physical injury and psychological problems, including loss of confidence and fear of recurrence, prolonged hospital stay, and financial burden [1,2]. Thus, the Joint Commission International set reducing risk of patient falls as the sixth of its International Patient Safety Goals [3].

The rate of falls in South Korean hospitals ranges from 0.08 to 0.19% [6,8]. However, these figures may be an underestimate considering that healthcare professionals are reluctant to report patient safety problems and hospitals tend not to disclose such statistics [2].

Globally, most accidental falls occur within the first week after admission [4,5,14], similar to what occurs in Korea [6]. However, most studies of falls occurring in a hospital setting have focused on identifying risk factors and developing predictive models. Few studies have identified factors affecting the time at which a fall occurs. Early falls can increase a patient's fear of remaining in the hospital and promote distrust of the hospital or healthcare professionals. Thus, risk factors related to early falls should be explored to prevent or reduce their occurrence [5].

Many hospitals have adopted electronic health record (EHR) systems. In Korea, 50.2% of tertiary teaching hospitals and 35.0% of general hospitals have adopted EHR systems [7]. As

a result, it is possible for researchers to use EHR data for patient care and outcome research.

However, the types of EHR data vary, from structured data, including vital sign data (e.g., blood pressure, heart rate, and respiratory rate) and drug prescription data, to unstructured data, including imaging and clinical narrative data [10]. In addition, a variable can be documented in multiple forms by various healthcare professionals and stored in multiple locations; thus, it is difficult to select or integrate EHR data.

Most studies of falls [2,5,6,8,12] considered only subjects who were registered as fallers in an adverse event reporting system (AERS). A previous study [9] extracted fall-related factors from narrative nursing records. Such records, in which nurses document patients' problems and the actions taken to solve them, can be used as a data source for fall events.

Based on this information, the present study identified patient characteristics and factors affecting the time at which a fall occurred by integrating and normalizing various EHR data from different sources.

Research Questions

The main research questions were:

1. When do patients fall after admission?
2. Do patient's or environmental characteristics influence timing of a fall?
3. What factors affect the time at which a fall occurs?

Methods

Setting and study subjects

This study was conducted in the neurology/neurosurgery and hematology/oncology wards of a Korean tertiary teaching hospital because that is where falls occur most frequently. This hospital has been using a standard terminology-based nursing records system since 2003 [15]. Fallers were extracted from the EHRs of those who were admitted to the hospital from January 2015 to May 2016.

In this study, a faller was defined as a patient who was registered in the AERS or a patient with a word or phrase such as "fall", "collapsed", "slip and fall", "fall off the bed", "fling oneself to the floor", "fall down during a move", "slip down during a move", or "flopped down" in their narrative nursing record.

Only the first fall was included for patients who fell more than once during a hospital stay.

Ethical Considerations

This study received approval from the institutional review board of the study hospital.

Procedures

Exploring fall-related variables

Fall-related variables were identified by reviewing clinical practice guidelines published by the World Health Organization, Agency for Healthcare Research and Quality, National Institute for Health and Care Excellence, Registered Nurses' Association of Ontario, and other organizations; fall risk assessment tools; and literature obtained from Google Scholar, PubMed, and the Cochrane Central Register of Controlled Trials.

Extraction of EHR data

We surveyed seven clinical nurses and two informatics specialists working at the study hospital to locate fall-related variables in the EHR system.

Variables were extracted from ten different data sources: clinical observation sheets, initial nursing assessments, the patient acuity classification system, laboratory results reports, doctor's medication orders, fall risk assessment forms, nursing activity sheets, surgical information sheets, doctor's progress notes, and narrative nursing records.

Data normalization

Data extracted from more than one source were normalized based on completeness, currency, and granularity of the record. "Completeness" refers to how often a variable is documented in a data source.

If variable A is documented more often in the narrative nursing record than on the clinical observation sheet, the narrative nursing record has higher completeness than the clinical observation sheet.

"Currency" refers to how close a variable is documented to the time of a fall to reflect the patient's status at the time of the fall. We defined the maximum time interval allowed between the recorded time and fall occurrence for each variable.

"Granularity" refers to the level of detail of information implied by the data. High granularity means a variable is documented with more attributes or with a detailed value set.

For example, "pain" data stored on both the clinical observation sheet and narrative nursing record were normalized (Figure 1). We first reviewed the clinical observation sheet where pain was documented more often than in the narrative nursing record. We assessed whether pain was documented in the clinical observation sheet within 2 days (the maximum time interval allowed) from a fall occurrence. If pain data recorded within 2 days from a fall occurrence was stored on the clinical observation sheet, we took the numeric rating scale (NRS) of pain. If not, we reviewed the narrative nursing record. If pain was documented in the narrative nursing record, we took the pain value (either the NRS score or Y/N). If pain was not documented in the narrative nursing record, then we concluded that the patient had no pain.

Statistical analyses

The characteristics of patients who sustained a fall and the time at which a fall occurred were summarized using descriptive statistics, including the mean, standard deviation (SD), median, interquartile range (IQR) and frequency. Second, univariate analyses were used to select significant variables affecting the time at which a fall occurred. Simple linear regression analyses were used for continuous variables such as age, patient acuity, and the fall-risk score. The t- or Mann-Whitney U test, a one-way analysis of variance, or the Kruskal-Wallis test were used for discrete variables such as sex, presence of a tube, and mental status, depending on the number of categories of independent variables and normality of the time-to-fall data. Statistically significant variables with a significance level of $p < 0.10$ were selected as independent variables for the multiple linear regression. A multiple linear regression analysis with backward variable selection was used to identify factors affecting the time at which a fall occurred. Variables with $p < 0.05$ were considered statistically significant. The time to fall from admission was natural logarithm-transformed because it was positively skewed. All analyses were performed using SPSS statistical software (ver. 23.0; IBM Corp., Armonk, NY, USA).

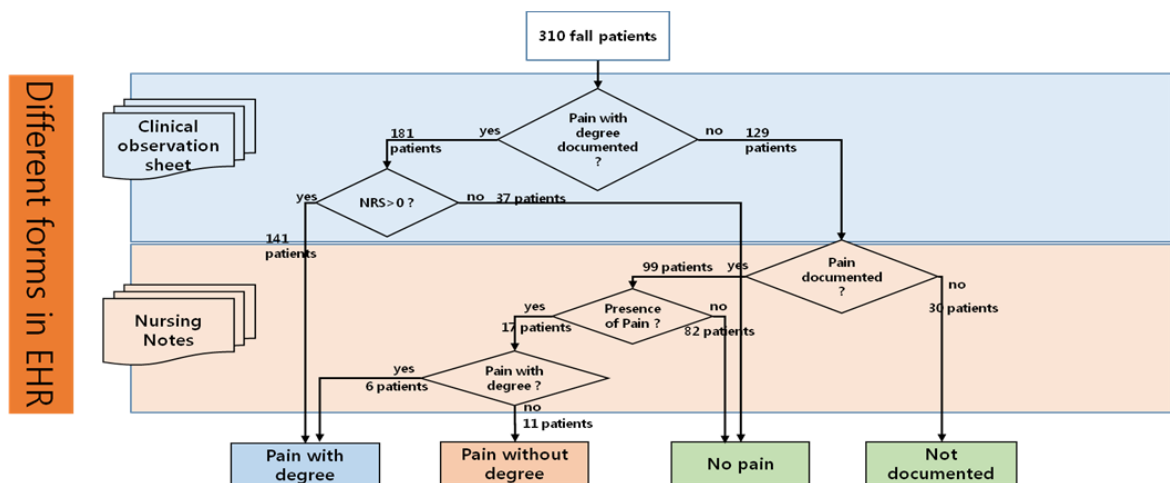


Figure 1 – Normalization process for the pain data

Results

Subject characteristics

A total of 310 patients who sustained a fall were included in the study (Table 1). The mean (SD) age was 62.9 (\pm 14.7) years and 186 (60.0%) patients were male. There were 151 (48.7%) patients who fell down within 1 week after admission. The median time to fall from admission was 7 days and the IQR was 11 (3.0–14.0) days. The mean (SD) and median (IQR) length of stay were 23.0 (\pm 19.2) days and 19 (20.7, 9.3–30.3) days, respectively.

We identified 205 (66.1%) fallers from the AERS and an additional 105 (33.9%) from narrative nursing records.

Figure 2 shows the distribution of the number of fallers by hospital days. The number of falls decreased as the hospital days increased.

Table 1 – Subject characteristics (N, number; M, mean; SD, standard deviation; IQR, interquartile range; AERS, adverse event reporting system)

Characteristics	Value	N (%) or M \pm SD or median (IQR)
Sex	Male	186 (60.0)
	Female	124 (40.0)
Age		62.9 \pm 14.7
	Neurology/ neurosurgery	147 (47.4)
Ward	Hematology/ oncology	163 (52.6)
Time to fall from admission		7 (3.0–14.0)
Length of stay		23.0 \pm 19.2
Fallers registered in the AERS		205 (66.1)
Fallers documented in the narrative nursing records		105 (33.9)

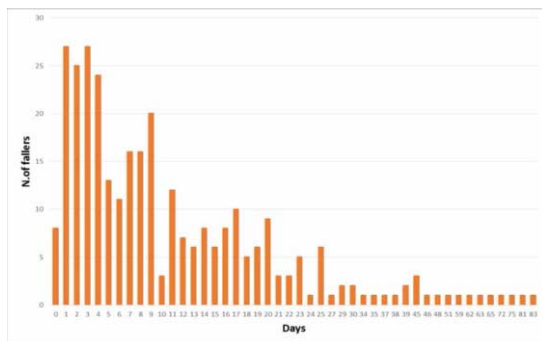


Figure 2 – Distribution of falls by hospital days

Extraction of EHR data by fall-related variables

In total, 95 variables were identified as factors related to falls through a literature review. They included patient variables (e.g., sex, age, education, and comorbidities), environmental variables (e.g., inadequate light and slippery floors), medical interventions (e.g., surgical procedure, drug administration, and non-invasive procedure), and nursing interventions (e.g., education for fall prevention, use of a bed rail, use of restraints, and a fall risk assessment).

Table 2 presents the types of data sources by variable. Data on “intravenous (IV) therapy” were extracted from clinical observation sheets, nursing activity sheets, the patient acuity classification system, and narrative nursing records. Data on “depression”, “confusion/disorientation”, and “altered elimination” were extracted from clinical observation sheets, fall risk assessment forms, and narrative nursing records. In addition, data on “pain”, “anxiety”, “hyperthermia”, “communication disorder”, “dizziness/vertigo”, and “enema” were extracted from two different sources, including the narrative nursing records. Of the ten data sources, the narrative nursing records had the highest coverage rates, with 88.2% of the variables stored. Variables extracted from more than one source were normalized for further analysis.

Table 2 – Possible types of data sources by variable (A, clinical observation sheet; B, initial nursing assessment; C, patient acuity classification system; D, laboratory results report; E, doctor’s medication order; F, fall risk assessment form; G, nursing activity sheet; H, surgical information sheet; I, doctor’s progress note; J, narrative nursing records)

Variables (example)	Type of data source									
	A	B	C	D	E	F	G	H	I	J
IV therapy	✓		✓				✓			✓
Depressed, confusion, altered elimination	✓					✓				✓
Tube/drainage (by type), restraints, bed rail	✓						✓			✓
Pain, anxiety, hyperthermia, communication disorder	✓									✓
Dizziness/vertigo						✓				✓
Enema			✓							✓
Age, sex		✓								
Fall-risk score						✓				
Drug (by type)						✓				
Operation									✓	
Hematocrit				✓						
Diagnosis									✓	
Walking aids, anorexia, fatigue, caregiver										✓

Relationship between the time to fall and variables

Of the 95 variables, 37 were identified as significant ($p < 0.10$) variables related to the time at which a fall occurred (Table 3).

Age, fall-risk score, vasodilators, antihistamines, anxiety, balance disorders, IV therapy, comorbid disease, hearing defects, an arterial line, and walking disorders were negatively related to the time to fall. That is, a patient with the above- mentioned variables tended to fall early.

Table 3 –Significant variables affecting the time to fall

Type of effect	Independent variables
Shortened time to fall (11)	Age, fall-risk score, vasodilator, antihistamine, anxiety, balance disorder, IV therapy, comorbid disease, hearing defects, arterial line, walking disorder
Delayed time to fall (26)	Education level, narcotic analgesics, anti- epileptics, selective serotonin reuptake inhibitor (SSRI), pain, benzodiazepine, antipsychotics, beta-blocker, calcium channel blocker, renin- angiotensin, anti-diabetics, diuretics, operation, urinary disorder, fall education, sleep disorder, dyschezia, anorexia, walking aids, lower sensory/motor disorder, fatigue, cognitive disorder, caregiver, central line, other drainage retention

Factors affecting the time to fall from admission

The assumptions of the linear regression such as autocorrelation (Durbin-Watson statistic = 2.151), multicollinearity (variance inflation factor value < 2.0), homoscedasticity, and normality of residuals were met.

The results of the multiple linear regression analysis between 37 independent variables and the time to fall are presented in Table 4. In total, 16 variables were significant.

Variables negatively related to time to fall included walking disorder, comorbid disease, IV therapy, and arterial line. These variables contributed to early falls after controlling for other variables. Variables positively related to time to fall included education level, fall prevention education, walking aids, caregiver, narcotic analgesics, SSRI, anti-diabetics, lower sensory disorder, cognitive disorder, dyschezia, anorexia, and central venous line. These factors contributed to delaying the time to fall after controlling for other variables.

The regression model was statistically significant ($F = 13.220$, $p = 0.000$). The adjusted R^2 was 0.458, with 16 independent variables explaining 45.8% of the variation in the time to fall.

Table 4 – Results of the multiple linear regression

Variables	β	SE	t	p
Constant	0.963	0.223	4.324	0.000
Education level: middle school	0.263	0.149	1.764	0.079
Education level: high school	0.335	0.125	2.673	0.008
Education level: \geq university	0.342	0.119	2.866	0.005
Narcotic analgesics	0.337	0.100	3.373	0.001
SSRIs	0.448	0.155	2.902	0.004
Anti-diabetics	0.421	0.111	3.780	0.000
Fall education	0.432	0.142	3.039	0.003
Walking aids	0.496	0.158	3.143	0.002
Lower sensory disorder	0.324	0.127	2.542	0.012
IV therapy	-0.230	0.095	-2.419	0.016
Comorbid disease	-0.508	0.110	-4.623	0.000
Cognitive disorder	0.678	0.219	3.090	0.002
Dyschezia	0.395	0.098	4.014	0.000
Anorexia	0.369	0.117	3.152	0.002
Caregiver	0.440	0.148	2.977	0.003
Arterial line	-0.899	0.311	-2.891	0.004
Central line	0.320	0.119	2.686	0.008
Walking disorder	-0.665	0.142	-4.697	0.000

$R^2=0.496$ Adjusted $R^2=0.458$ $F=13.220$ $p=0.000$

Discussion

Factors affecting the time to fall from admission were explored using EHR data. Previous studies of falls in a hospital setting extracted fall patients only from an AERS [1,2,5,6,8]. However, in this study, we identified 105 additional fall patients by analyzing narrative nursing records, which were a useful complement to the AERS to identify fall occurrences. Using various EHR data sources, we were able to extract more independent variables, including presence of caregivers and whether patients had been educated in fall prevention.

EHR data are stored on multiple databases in a variety of data types by different users. Therefore, it is necessary for researchers to integrate scattered, heterogeneous data from EHRs before analyses [10].

Data quality should be assured during data integration. The important attributes of data quality are “accuracy” and “completeness” [11]. Data accuracy is the extent to which documented data match the truth, and data completeness refers to the extent to which data have actually been documented [11]. In this study, we applied normalization criteria based on

completeness, currency, and granularity to integrate the EHR data and assure data quality. These criteria helped to improve data accuracy and completeness.

We reviewed the completeness of ten different data sources by examining how frequently data were stored at each source to determine which source to use first. Of the ten data sources, the narrative nursing records had the highest completeness for 58 (61.1%) variables. We extracted these variables from the narrative nursing records first.

Once the order of sources to use first was determined in terms of completeness, we selected data as current as possible from the source to reflect the patient’s state at the time of the fall. We defined the maximum time interval allowed between data extraction and the fall event for each variable. If variable A had a property that changed from time to time, we set a shorter maximum time interval for variable A. Mental state, including confusion or drowsiness, pain, and tube/drainage retention, were included in this category. If there was no record of a variable within the maximum time interval allowed, we assumed that the patient did not have the variable. For example, if there were no data on “confusion” or “unconsciousness” within 1–2 days before a fall, we regarded the patient’s consciousness as alert/awake. However, demographic variables, including age, sex, education, and comorbidity, were extracted from data recorded at the time of admission, as these variables do not change.

After we collected the most current data, we reviewed the granularity of the data. For example, pain, consciousness, muscle strength, and communication disorders documented in the clinical observation sheet are assessed using rating scales (the NRS, Glasgow Coma Scale, motor power, and NIH Stroke Scale, respectively). The degree of a variable could also be extracted from the narrative nursing record. For example, narrative nursing statements such as “patient is unconscious”, “patient is confused”, and “patient is fully awake” express different states of a patient’s consciousness. If it was impossible for us to determine the degree of a variable, we categorized the variable with a Boolean (Y/N) value.

Data values for the same variable are stored inconsistently at different data sources in terms of data type, so we tried to reduce this inconsistency by normalization. Missing values were reduced by this process. For example, we reduced the number of missing values in the pain data by 29% using clinical observation sheets and narrative nursing records during the normalization process, which improved the external validity of the results.

The greatest number of falls occurred during the first week of the hospital stay. This finding is similar to that of previous studies [4,5,6,14]. This could be because newly admitted patients may have experienced difficulties adjusting to the new environment [14]; they were not used to the bed, toilet, or medical devices. Therefore, it is important to assess a patient’s fall risk within the first week after admission. Patients with a high risk of falling during the first week of admission deserve special attention from healthcare professionals [5].

Walking disorders, comorbid disease, IV therapy, and arterial lines were negatively related to the time to fall from admission. Walking disorders, comorbid disease, and IV therapy are components of the Morse Fall Scale, which is used to assess a patient’s likelihood of falling. They are not only risk factors for falling, but also factors affecting the time to fall according to this study. Patients with a walking disorder tended to fall early, as most of these patients were hospitalized in the neurology/neurosurgery ward with poor mobility and walking disturbances due to their disease (e.g., stroke or

Parkinson's disease). This finding is similar to that of previous studies reporting that an unsafe gait is a risk factor contributing to the time to fall [5] and the occurrence of falls [12]. Comorbid diseases such as hypertension, diabetes mellitus, hyperlipidemia, cardiac disease, and liver disease are related to an increased risk of inappropriately prescribing drugs and adverse drug reactions [13], which may cause an early fall. In addition, patients receiving IV therapy or those with an arterial line are more likely to fall by tripping on the IV poles or related equipment.

The presence of a caregiver and use of walking aids were significant variables for delaying the time to fall. A caregiver and walking aids help patients maintain their balance when they walk. Fall prevention education by nurses also prolongs the time to fall. Therefore, early education for fall prevention is essential to reduce the risk of an early fall.

This study has some limitations. First, we used a single research site and selected the neurology/neurosurgery and hematology/oncology wards (where falls occurred most frequently in the hospital). Therefore, there are constraints on generalizability and applications to practice. Second, environmental variables such as inadequate light, a slippery floor, and unfitted shoes were not included in the analyses, as they were not documented in the EHR. In this study, we only examined the time to fall after admission of the fallers. Currently, we are working on a model to predict the risk of falling with faller and non-faller data.

Conclusion

We identified factors affecting the time to fall using EHR data. We used ten different data sources to extract variables, and the data were integrated and normalized based on normalization criteria such as completeness, currency, and granularity. This process helped improve the accuracy and completeness of the data. In this study, the narrative nursing records were very useful for identifying fall patients and fall-related variables; they also complemented the AERS.

This study shows that the greatest number of falls occurred during the first week of a hospital stay. In particular, a walking disorder, comorbid disease, IV therapy, and arterial line were related to an early fall. We recommend that nurses assess early fall risk and pay extra attention to patients during the initial period of hospitalization.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea government (NRF-2015R1A2A2A01008207) and Seoul National University Big Data Institute through the Data Science Research Project 2016.

The English in this paper has been checked by at least two professional editors, both native speakers of English. For a certificate, please see: <http://www.textcheck.com/certificate/hjf9Fs>

References

- [1] T.J. Dunne, I. Gaboury, and M.C. Ashe, Falls in hospital increase length of stay regardless of degree of harm, *J Eval Clin Pract* **20**(4) (2014), 396-400.
- [2] H.J. Hong, N.C. Kim, Y. Jin, J. Piao, and S.M. Lee, Trigger factors and outcomes of falls among Korean hospitalized patients: analysis of electronic medical records, *Clin Nurs Res* **24**(1) (2015), 51-72.

- [3] Joint Commission International, Joint Commission International Accreditation Standards for Hospital, 5th Edition, Joint Commission International, 2013.
- [4] G.M. Tinker, Accidents in a geriatric department, *Age Ageing* **8**(3) (1979), 196-198.
- [5] M. Vassallo, J.C. Sharma, R.S. Briggs, and S.C. Allen, Characteristics of early fallers on elderly patient rehabilitation wards, *Age Ageing* **32**(3) (2003), 338-342.
- [6] Y.S. Kim, and S. Choi-Kwon, Fall risk factors and fall risk assessment of inpatients, *Korean J Adult Nurs* **25**(1) (2013), 74-82.
- [7] D. Yoon, B.C. Chang, S.W. Kang, H. Bae, and R.W. Park, Adoption of electronic health records in Korean tertiary teaching and general hospital, *Int J Med Inform* **81**(3) (2012), 196-203.
- [8] C.G. Kim, and M.J. Suh, An analysis of fall incidence rate and its related factors of fall in inpatients, *Journal of Korean Society of Quality Assurance in Health Care* **9**(2) (2012), 210-228.
- [9] I. Cho, I.S. Park, and E.M. Kim, Exploring the utility of the ICNP based electronic nursing records as a research source for inpatients' fall, *Perspectives in Nursing Science* **5** (2008), 33-43.
- [10] P.B. Jensen, L.J. Jensen, and S. Brunak, Mining electronic health records: towards better research applications and clinical care, *Nat Rev Genet* **13**(6) (2012), 395-405.
- [11] D.G. Arts, N.F. De Keizer, and G.J. Scheffer, Defining and improving data quality in medical registries: a literature review, case study, and generic framework, *JAMIA* **9**(6) (2002), 600-611.
- [12] S.H. Yoo, S.R. Kim, and Y.S. Shin, A prediction model of falls for patients with neurological disorder in acute care hospital, *J Neurol Sci* **356**(1) (2015), 113-117.
- [13] G.E. Caughey, E.E. Roughead, S. Shakib, R.A. McDermott, A.I. Vitry, and A.L. Gilbert, Comorbidity of chronic disease and potential treatment conflicts in older people dispensed antidepressants, *Age Ageing* **39**(4) (2010), 488-494.
- [14] K. Rapp, J. Ravindren, C. Becker, U. Lindemann, A. Jaensch, and J. Klenk, Fall risk as a function of time after admission to sub-acute geriatric hospital units, *BMC Geriatr* **16**(1) (2016), 173.
- [15] I. Cho, Assessing the quality of structured data entry for the secondary use of electronic medical records, *J Kor Soc Med Informatics* **15**(4) (2009), 423-431.

Address for correspondence

Hyeoun-Ae Park, PhD

College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 110-799, South Korea.

Tel: +82-2-7408827, Fax: +82-2-765-4103,

E-mail: hapark@snu.ac.kr

Identifying and Synchronizing Health Information Technology (HIT) Events from FDA Medical Device Reports

Hong Kang^a, Frank Wang^b, Sicheng Zhou^a, Qi Miao^c, Yang Gong^a

^aSchool of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA,

^bCollege of Natural Sciences, The University of Texas at Austin, Austin, TX, USA,

^cSchool of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

Abstract

Health information technology (HIT) events, a subtype of patient safety events, pose a major threat and barrier toward a safer healthcare system. It is crucial to gain a better understanding of the nature of the errors and adverse events caused by current HIT systems. The scarcity of HIT event-exclusive databases and event reporting systems indicates the challenge of identifying the HIT events from existing resources. FDA Manufacturer and User Facility Device Experience (MAUDE) database is a potential resource for HIT events. However, the low proportion and the rapid evolvement of HIT-related events present challenges for distinguishing them from other equipment failures and hazards. We proposed a strategy to identify and synchronize HIT events from MAUDE by using a filter based on structured features and classifiers based on unstructured features. The strategy will help us develop and grow an HIT event-exclusive database, keeping pace with updates to MAUDE toward shared learning.

Keywords:

Patient Safety; Medical Errors; Information Storage and Retrieval

Introduction

Health information technology, or Health IT (HIT) -including electronic health records (EHR), computerized provider order entry (CPOE), clinical decision support (CDS) systems, and personal health records (PHR) - has been placed in the spotlight for its bold promises to increase hospital efficiency, improve patient safety, and reduce medical errors [1]. According to the Agency for Healthcare Research and Quality (AHRQ), HIT is defined as “the use of information and communication technology in healthcare to support the delivery of patient or population care or to support patient self-management” [2]. While the use of HIT presents many new opportunities to improve patient care and safety, it can also create new hazards and opportunities for error. HIT will fulfill its potential only if the risks associated with its use are identified and a coordinated effort is developed to mitigate those risks. In fact, one in every six patient safety events (PSE) can be attributed to HIT [3], making HIT one of the top 10 technology-related hazards as identified by the Emergency Care Research Institute [4].

Event reporting is a potential approach for shared learning, which has been proven in many high-risk industries, such as the aviation, nuclear, and railroad industries. For HIT to become more widely adopted and trusted, it is crucial to gain a better understanding of the nature of the errors and adverse events caused by current HIT systems. Identifying problems of HIT systems when they occur and presenting to stakeholders is an

area that has not received enough emphasis [5]. Similar to PSE reporting, an effective means suggested by the Institute of Medicine, acquiring knowledge from previous experiences to prevent the recurrence and serious consequences of similar HIT events could be a practical start [6]. PSE reporting is frequently a mainstay of frontline practitioners’ efforts to detect PSEs and quality problems [7]. The reports collected from a broad range of stakeholders can generate a summary and feedback toward actionable knowledge and shared learning.

Although AHRQ has taken initial steps to standardize HIT reporting by the creation of the Common Formats [8], currently there is still no HIT exclusive reporting form/system. Part of the reason may lie in the fact that current definitions of HIT are often broad with ill-defined borders [9]. Under the current AHRQ HIT definition, almost any medical device that utilizes electronic software and hardware can be considered an HIT device. In practice, however, reporters may find a stricter definition of HIT more helpful in determining the events involving technology and should be reported under HIT. The starting point for creating an effective HIT exclusive reporting form/system lies in the creation of an HIT event database. The database will allow researchers to analyze connections among HIT events, identify common themes of technology-induced errors, synchronize HIT related reports from existing resources, and develop a classification or terminology to standardize HIT reporting.

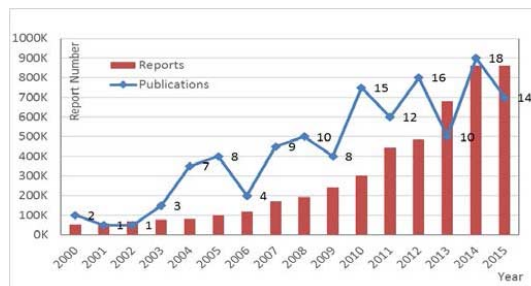


Figure 1– Trend of new MAUDE reports per year (bars) and MAUDE related publications per year (line) since 2000.

Search resources for peer reviewed publications included three databases: Ovid MEDLINE, PsycINFO, and Health and Psychological Instruments.

The scarcity of HIT event-exclusive databases and event reporting systems indicates the challenge of identifying and synchronizing the HIT events from existing resources. The FDA Manufacturer and User Facility Device Experience (MAUDE) database [10] is a rich and publicly accessible resource of HIT events. The MAUDE database contains the

reports of events involving medical devices, voluntary reports of medical device malfunction, and reports of problems leading to serious injury and death since June 1993. The MAUDE database houses medical device reporting submitted to the FDA by mandatory reporters (manufacturers, importers and device user facilities) and voluntary reporters (healthcare professionals, patients and consumers). MAUDE data are updated weekly and searchable online. As of November 2016, MAUDE had more than 5 million reports. Although the number of MAUDE-based HIT publications has increased in recent years (Figure 1), the number remains fractional given the enormous number of reports in MAUDE. Most publications were reviews about certain events, devices, or treatments based on MAUDE reports. Exploration of the utility of MAUDE for understanding HIT problems has been limited [11]. Therefore, a search and classification strategy holds potential for using the HIT reports and would result in a database to store, manage, and compare HIT reports and to identify and analyze HIT solutions [12].

In this study, we created a comprehensive search strategy that utilizes both structured (device data) and unstructured (text data) to extract HIT related events from the FDA MAUDE database. Our strategy first involved the use of a keyword filter on the structured data to screen the FDA database for reports related to HIT. Then machine learning algorithms were used to classify the selected reports based on their narrative text. Finally, reports classified with a high probability of HIT were added to our HIT event database. Using this strategy, we were able to grow an HIT event database composed of HIT related events from the FDA MAUDE database. Ultimately, the creation of this database holds promise in aiding the understanding, characterization, discovery, and reporting of HIT related events.

Methods

Construct a Filter for HIT Related Reports Based on Structured Data

The device data of FDA MAUDE consists of 45 structured fields that contain information regarding the device involved in the event. Most of the fields, however, are either left blank by reporters or are of little use for the purpose of identifying and synchronizing HIT related events. After reviewing all the fields, we found that the *generic name* and *manufacturer name* fields have the greatest potential in identifying HIT related events and thus we utilized both in the creation of our filter.

To develop the filter, we started with a set of common computer hardware and software related keywords that had been previously identified [13]. The starting keyword list was expanded by the addition of several generic terms such as “software,” “program,” and “hardware” and several more modern terms such as “electronic medical record” and “portal technology.” Then all of the generic names from the FDA MAUDE database starting from Jan 2010 to Dec 2015 were extracted, yielding a total of 60,000 unique generic names. Next, through partial keyword matching of the generic names to the keywords list, a subset of generic names appearing HIT-related was extracted. The subset was further analyzed by domain experts to determine which generic names were actually linked to HIT reports by using a small portion of the 2015 FDA MAUDE database.

A similar approach was utilized in selecting a list of manufacturers for the filter. We started with a list of seven popular HIT manufacturers [13] and then added 347

manufacturers of HIT software. The reports from the 2015 FDA MAUDE database related to those manufacturers were extracted and checked to determine the most relevant manufacturers.

Evaluate the HIT Filter Through Expert Review

The HIT filter was first applied on the 2015 FDA MAUDE database. Then a subset (10%) of the reports screened by the filter was selected for manual review by two domain experts. The experts labeled each report with one of three labels: HIT, Not HIT, or Unsure. The reports that the reviewers disagreed on or were unsure about were resolved through group discussion.

During reviewing, we narrowed the HIT definition to identify the most clinically relevant and consequential HIT devices. Under our current understanding, an HIT device is any device that utilizes both hardware and software to facilitate health information exchange in order to aid in the diagnosis, treatment, or prevention of disease. Using this definition, priority is given to HIT systems that focus on information exchange such as electronic health records, computerized physician order entry, and picture archiving communications systems. Implantable devices, glucose monitors, defibrillators, and similar devices are excluded under this definition as they do not actively facilitate health information exchange.

Construct HIT Classifiers Based on Unstructured Data

Machine learning classifiers including logistic regression, support vector machine (SVM), naïve Bayes, and random forest were constructed using the unstructured data (narrative text) of the reviewed reports. Each report in the labeled training set was treated as a vector of words and was weighted by a term frequency-inverse document frequency (TF-IDF) schema. Each classifier was evaluated using leave-one-out cross-validation (LOOCV) and performance was weighted based on both their F1 score and ROC curve. The best classifier was selected and then attribute selection was done to further increase the performance and efficiency of the classifier.

Grow an HIT Event Database

The optimized HIT classifier was applied to previous years of the FDA Maude database starting from 2015. The reports were ranked in order of their probability of being HIT related. The probability was calculated internally by the classifier. The ranked reports were then manually reviewed until less than 90% of the reports at a given probability threshold were HIT related. The reports above the probability threshold were then added to the rest of the HIT reports previously found, forming the final HIT event database.

Results

Keywords of the HIT Filter

A subset of 336 software and 749 hardware generic names most likely to be related to HIT was extracted from the 3,634,879 reports in MAUDE during 2010-2015, which account for 72% of the total reports in MAUDE since 1993. After the expert review, 58 keywords from *generic names* (39 software and 19 hardware keywords respectively) and 16 keywords from *manufacturer names* were determined to compose the filter for HIT related reports, as shown in Table 1.

Table 1– Keywords of the HIT filter in alphabetical order

58 keywords from generic names
Software (39) ADC, Alert, Automated Dispensing Cabinet, Communication Device, Communication System, CPOE, Data Backup, Database, Decision Support, Digital, Dispensing System, Dose Suggestion, Downloader, Drug Suggestion, EHR, Electronic Health, Electronic Medical, Electronic Patient, EMR, ICT, Imaging System, Information System, Internet, Invision, LIS, Management System, Monitoring System, Network, Order Entry, PACS, Picture Archiving, Portal, Powerchart, Program, Server, Soarian, Telemetry, Trima Accel Platelet, Web Hardware (19) Anesthesia Monitor, Apnea Monitor, Arterial Monitor, Atlas Monitor, Blood Pressure Monitor, Central Monitor, Computer, Console Monitor, Drug Screen, Fetal Monitor, Patient Monitor, Physiological Monitor, Pressure Monitor, PT Monitor, Safety Monitor, Telemetry Monitor, Telemetry Transmitter, Vital Sign Monitor, Workstation
16 keywords from manufacturer names
Allscripts, Centricity, Cerner, Epic, Ge Healthcare, Hass, Healthtronics, Henry Schein, Homer, Isite, iSOFT, Kestral, McKesson, Medical Director, MedPro, Oasis

The filter was first applied on the 2015 FDA MAUDE database including 860,915 reports. 4871 reports (2479 software and 2392 hardware reports) were initially found. 490 reports (10%) were randomly selected according to the keyword distribution for expert review and labeling. 312 reports were identified as HIT related by experts, which means the filter can generate a report subset from original MAUDE database with about 63.7% HIT related reports. This proportion is significantly higher than 0.1%, which is the estimated proportion of HIT related reports in the entire MAUDE database [12].

HIT Classifiers

TF-IDF models were applied on the narrative data to further identify HIT events after using the filter based on the structured data. We used the manually labeled reports (312 HIT and 178 non-HIT reports) from the 2015 MAUDE database to build the training set. The same number of non-HIT-related reports was randomly selected to compose the training data with the HIT-related reports. The classifiers were trained using four methods: logistic regression, random forest, naïve Bayes, and SVM. As shown in Table 2, logistic regression and SVM exhibit better performances among the four methods.

Table 2– HIT classifier performance

Methods	Accuracy	F-Measure	ROC
Logistic Regression	88.6%	0.885	0.919
Random Forest	83.5%	0.835	0.934
Naïve Bayes	84.8%	0.846	0.946
SVM	88.6%	0.886	0.886

The Growing HIT Event Database

Due to the enormous size of MAUDE and the small percentage of HIT events (about 0.1%), directly identifying and extracting all HIT events from MAUDE are almost impossible using a straightforward strategy. The classification of data with imbalanced class distribution has encountered a significant drawback of the performance attainable by most standard classifier learning algorithms [14]. The aim of this study is to establish a model which can identify and synchronize HIT related events from MAUDE database toward an HIT exclusive database for shared learning. Therefore, we need to keep the False Positive (FP) rate within a low value to make sure the quality of the database, even additional HIT reports are missing (high False Negative (FN) rate). Random forest model can help reduce FP rate by trading in FN rate. The strategy was setting a threshold to the confidence coefficient and only taking the samples whose prediction confidence coefficients were higher than the threshold. As shown in Figure 2, 0.7 could be an

appropriate threshold since more than 90% FP samples are excluded with a loss of less than a half HIT related reports.

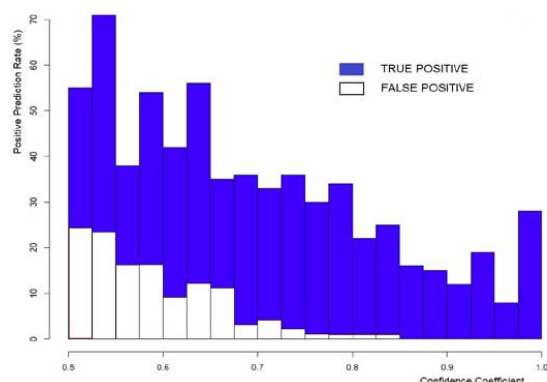


Figure 2– TP and FP distribution of the random forest model

To grow a database for HIT event reports and further improve the HIT filter and classifiers, we are applying the HIT event filter and classifiers on the 2014 dataset. Only the reports that are identified as HIT will be manually reviewed. All the reviewed reports will be added to the training set, which will consequently help build new classifiers. The same process will, in turn, be retrospectively applied on the datasets from 2013, 2012, 2011, 2010, and beyond (Figure 3).

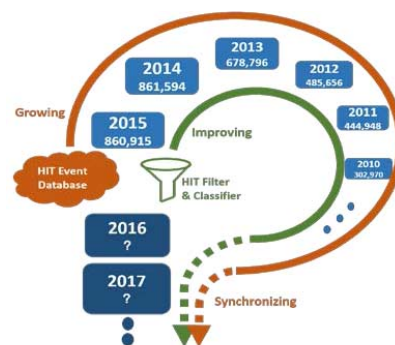


Figure 3– The strategy of improving the HIT filter and classifier and growing the HIT event database

At the end of each iteration, the four training methods will be re-evaluated based on the corresponding manual review, and the best method will be used in the classifier. The classifier is expected to be improved as more labeled reports are included. As the manually reviewed HIT events are accumulated, an HIT event database will be established and keep growing. This mechanism will keep synchronizing HIT related events from MAUDE database to the HIT database in the future.

Discussion

Offer a Broader and Organized View of HIT Events

Some HIT events may seem trivial but could represent much larger and more important problems, which is similar to a tip of a very large and dangerous iceberg. We developed a mechanism of identifying and synchronizing HIT related events from FDA MAUDE database. The outcome is a timely reflection of the evolvement of HIT events and is helpful for enriching HIT

knowledge and better using the historic reports toward an overall understanding and analysis of the characteristics, occurrence, observation, and description of HIT events. Using our proposed filter and classifiers, fragmented and isolated HIT events could be better understood when the connection with other relevant events are offered.

Challenges of Identifying HIT Events

While machine learning has been employed successfully in many contexts such as spam filtering and social media sentiment analysis, its application to HIT event identification is still challenging. Early methods of HIT event identification relied on flagging cases that contained certain technology related keywords for further human review. While these approaches were feasible on a small scale, the growing use of technology in healthcare has led to a rapid rise in technology related events. Meanwhile, the events involving technology have grown in complexity, making researchers sometimes feel difficult to decipher whether or not the technology involved was truly an integral part of the event. Although various natural language processing methods exist such as N-grams and concordance to aid in textual understanding, the classification of PSEs based solely on narrative text may still be out of the reach of current machine learning techniques.

As a result, we utilized both the structured and unstructured data of the FDA MAUDE database. In the device data reports, the two most useful fields for creating our HIT filter were generic and manufacturer name. The generic name field was especially informative as it allowed for the quick identification of device concerning each report. In some cases, the generic name field was pivotal to understanding the event as the device of interest could not be deduced from the narrative reports alone. The manufacturer name field is also proven useful as in some cases, the generic name may have referred to a new or more sophisticated product that was not yet recognized by our list of previously identified generic names. By utilizing both fields, the proposed HIT filter was able to capture more potential HIT events than it would have if only the generic name field were utilized.

Quickly Changing Nature of HIT

Each year new generic names are created and some older names fall into disuse in MAUDE. The number of generic names has shown a growing trend, with each year starting from 2010 containing approximately 1000 more terms than the previous year. This challenge was addressed during the selection of our generic name keywords by reducing the complexity of the generic names while still allowing them to retain their specific meanings. As an example, if the generic name originally found in the database was “picture archiving communications system,” the keyword added to our final list was “picture archiving.” By reducing the complexity of the keywords in our filter, the ability to capture new variations of the same generic name in the future was greatly improved. Furthermore, as previously mentioned, the manufacturer name field was used to supplement the generic names. Cases with new and unrecognized devices could still be detected by the HIT filter if they were made by a recognized HIT manufacturer.

Classifier Error Analysis and Optimization

Two domain experts independently reviewed the filtered 2015 HIT reports and had an inter agreement kappa score of 0.9, which was used as a proxy for the maximum performance that the classifier could reasonably achieve. The initial classifier tended to make more false negatives (true HIT cases labeled as non-HIT) than false positives (non-HIT cases labeled as HIT)

in error analysis. In improving the classifier, greater emphasis was placed on reducing the false negatives than the false positives. This is because in practice, reports that are falsely labeled as being HIT can usually be disregarded by users of the database with little effort. Conversely, true HIT reports labeled as non-HIT may pose a much greater risk as users may be unable to learn from those events. By using the proposed approach, while our classifier’s overall accuracy may have been reduced, the classifier we built has a greater ability to capture new and unique HIT safety events.

Evaluation of Proposed HIT Extraction and Synchronization Method

The method we have developed for extracting and synchronizing HIT reports from the FDA MAUDE database shows promise in staying up to date with future changes in HIT. As new devices and technologies are emerging every year, HIT event database faces the risk of quickly becoming outdated. By utilizing an easily updatable device data filter, new types of HIT can be identified and added to the existing database, providing users with the most up-to-date information on HIT. Furthermore, the text classifier we have built allows for the timely classification of HIT related events, greatly reducing the need for future human labor to maintain the database. With an up-to-date and comprehensive HIT event database, researchers may be able to gain a deeper understanding of the nature of HIT related events and their potential consequences. As well, the database may aid in the classification of HIT related events and eventually lead to a universally accepted HIT exclusive reporting system.

Importance of the HIT Event Database

In a high stakes field such as healthcare, it is critical that HIT events are reported and that manufacturers are held accountable for their products. However, manufacturers may not have enough resources allocated towards diagnosing and fixing those issues even when the events are reported. While the staggering pace of technology has driven much innovation in healthcare, the need to take a careful look at the HIT related events has never been greater. The proposed HIT event database offers an opportunity to compare, analyze, and integrate similar HIT events, and ultimately keep HIT on the right track towards becoming a safe and integral part of our healthcare system.

Limitations

The FDA MAUDE database is currently the only publicly accessible resource to contain HIT events. The database is built on voluntary reports and utilizes only a passive surveillance system to verify reports. Consequently, the database may contain incomplete, inaccurate, or biased reports, preventing conclusions regarding the frequency or prevalence of events from being drawn. Nevertheless, the database still contains highly informative data and can be used to better understand the nature and scope of different PSEs.

Another limitation lies in the initial selection of keywords that were used to later find common generic names of HIT devices. While the initial keyword list contained a comprehensive amount of computer and software related terms, it may not have included all possible words that could be used to describe HIT related devices. As a result, few HIT events that were described in an uncommon manner may have been missed during our initial search. In the future, the proposed HIT event database may help overcome this challenge by aiding in the development of a standardized reporting terminology for HIT events. With a standardized reporting terminology, it will likely be easier for

reporters to describe HIT events in a manner that is more useful and informative to researchers.

Future work

Improve Classifiers by Using Semi-supervised Learning

Manually reviewing all cases in the FDA MAUDE database was simply infeasible and that machine learning was likely to be the only viable approach. Traditionally, the two paradigms of machine learning have been supervised (all labeled data) and unsupervised learning (all unlabeled data). However, much attention has been recently placed on semi-supervised learning for its ability to utilize only a small amount of labeled cases combined with a large amount of unlabeled cases to improve classification accuracy. In the case of HIT, this approach seems well-suited as the cost of labeling narrative text by manual review is quite high, while the cost of obtaining unlabeled reports is minuscule in comparison. One of the simplest methods within semi-supervised learning is self training. In creating our HIT event database, we will utilize a method of iterative self-training to grow the database in a timely manner. It will become a viable method to extract and classify HIT reports from large databases such as the FDA MAUDE.

Utilize the HIT Event Database

More work needs to be done to understand, organize, and uncover the relationships among the HIT reports contained in the database. To fully utilize the database, the themes and topics of the reports must be mapped and understood in context of one and another. Doing so will allow for a better understanding of the relations among HIT events and will greatly aid in the understanding of how HIT changes over time. One potential tool for topic modeling is Latent Dirichlet Allocation (LDA). This commonly used algorithm can be used to analyze unannotated text and discover shared themes among reports. After common themes of HIT related events are found, experts may be able to focus their efforts on addressing these issues and develop actionable solutions to prevent and minimize the risks that patients face during HIT related events. Ultimately, this will advance HIT a safe and integral part of healthcare and improve patient safety.

Conclusion

We proposed a strategy to identify and synchronize HIT events from FDA medical device reports, and to grow an HIT event-exclusive database. The database provides a resource for stakeholders to analyze connections among HIT events, identify common themes of technology-induced errors, synchronize HIT related reports from existing resources, and develop a classification or terminology to standardize HIT reporting toward actionable knowledge and shared learning.

Acknowledgements

This project is supported by UTHealth Innovation for Cancer Prevention Research Training Program Post-Doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant #RP160015), Agency for Healthcare Research & Quality (1R01HS022895), and University of Texas System Grants Program (#156374).

References

- [1] C.R. Denham, D.C. Classen, S.J. Swenson, M.J. Henderson, T. Zeltner, and D.W. Bates, Safe use of electronic health records and health information technology systems: trust but verify, *J Patient Saf* **9** (2013), 177-189.
- [2] AHRQ Announces Interest in Research on Health IT Safety (Notice Number: NOT-HS-16-009), in, Agency for Healthcare Research and Quality, March 9, 2016.
- [3] K.C. Cheung, W. van der Veen, M.L. Bouvy, M. Wensing, P.M. van den Bemt, and P.A. de Smet, Classification of medication incidents associated with information technology, *J Am Med Inform Assoc* **21** (2014), e63-70.
- [4] Top 10 health technology hazards for 2015, in, Emergency Care Research Institute (ECRI), November 2014.
- [5] Identification and prioritization of health IT patient safety measures., in, Washington, DC, February 2016.
- [6] W.E. Fassett, Patient safety and quality improvement act of 2005, *Annals of Pharmacotherapy* **40** (2006), 917-924.
- [7] Voluntary patient safety event reporting (incident reporting), in, Agency for Healthcare Research and Quality, August 2014.
- [8] Common Formats, in, Agency for Healthcare Research and Quality.
- [9] D.F. Sittig and H. Singh, Defining health information technology-related errors: new developments since to err is human, *Arch Intern Med* **171** (2011), 1281-1284.
- [10] MAUDE - Manufacturer and User Facility Device Experience, in, U.S. Food and Drug Administration.
- [11] R.B. Myers, S.L. Jones, and D.F. Sittig, Review of Reported Clinical Information System Adverse Events in US Food and Drug Administration Databases, *Appl Clin Inform* **2** (2011), 63-74.
- [12] F. Magrabi, M.S. Ong, W. Runciman, and E. Coiera, Using FDA reports to inform a classification for health information technology safety problems, *J Am Med Inform Assoc* **19** (2012), 45-53.
- [13] F. Magrabi, M.S. Ong, W. Runciman, and E. Coiera, An analysis of computer-related patient safety incidents to inform the development of a classification, *J Am Med Inform Assoc* **17** (2010), 663-670.
- [14] Y. Sun, A.K.C. Wong, and M.S. Kamel, Classification of imbalanced data: a review, *International Journal of Pattern Recognition and Artificial Intelligence* **23** (2009), 687.

Address for Correspondence

Yang Gong, MD, PhD

7000 Fannin St Suite 600, Houston, TX 77030, USA

Yang.Gong@uth.tmc.edu

Phone: +1 713 500 3547, Fax: +1 713 500 3907.

Comparative Performance Analysis of Different Fingerprint Biometric Scanners for Patient Matching

Noah Kasiiti^a, Judy Wawira^b, Saptarshi Purkayastha^c, Martin C. Were^{a,d,e}

^a Institute of Biomedical Informatics, Moi University, Eldoret, Kenya

^b Department of Radiology, Indiana University, Indianapolis, IN, USA

^c School of Informatics, Indiana University-Purdue University Indianapolis, Indianapolis, IN, USA

^d Department of Biomedical Informatics & Medicine, Vanderbilt University, Nashville, TN, USA

^e Vanderbilt Institute of Global Health, Vanderbilt University, Nashville, TN, USA

Abstract

Unique patient identification within health services is an operational challenge in healthcare settings. Use of key identifiers, such as patient names, hospital identification numbers, national ID, and birth date are often inadequate for ensuring unique patient identification. In addition, approximate string comparator algorithms, such as distance-based algorithms, have proven suboptimal for improving patient matching, especially in low-resource settings. Biometric approaches may improve unique patient identification. However, before implementing the technology in a given setting, such as health care, the right scanners should be rigorously tested to identify an optimal package for the implementation. This study aimed to investigate the effects of factors such as resolution, template size, and scan capture area on the matching performance of different fingerprint scanners for use within health care settings. Performance analysis of eight different scanners was tested using the demo application distributed as part of the Neurotech Verifinger SDK 6.0.

Keywords:

Biometry; Patient Identification Systems; Health Information Systems

Introduction

A critical component of health care delivery is the ability to correctly identify the individual receiving care and access their medical record. Failure to correctly match patients is a major contributor to inefficiencies in care delivery and medical errors [1]. For example, in the United States, about 195,000 deaths occur each year because of medical errors, with 10 of 17 being the result of identity errors [2]. The problem is even worse in low and middle income countries (LMICs) where very few accurate and comprehensive person identification procedures guaranteeing the unambiguous identification of their citizens from the day they are born, have been implemented [3]. Poor national person identification systems, inefficient identification procedures, and the use of weak search criteria further aggravate the problem.

In the western world, researchers and hospital administrators largely rely on deterministic or probabilistic algorithms and other statistical matching procedures for patient identity management [4]. Deterministic matching algorithms use an exact match or rely on comparisons between two fields [5]. As an example, deterministic matching can be used to compare unique identifiers, such as national IDs, to determine a match.

Probabilistic matching, which is by far the most widely implemented technique for record matching, does not depend on unique identifiers. For probabilistic matching, the values between two records are compared across several fields and weights are assigned based on how close the values in the corresponding fields are. The sum of these weights indicate the closeness of the match for the compared records [4-6]. In the case of patient identification, probabilistic matching would compare closeness using key patient identifiers such as name, address, national IDs, and date of birth.

The application of these statistical matching approaches is limited, especially in developing countries. Deterministic matching in these settings is limited when no single field can provide a reliable match between two records. In many cases, the order that a patient's first, middle, and last name are recorded differs between visits, addresses are unreliable, dates of birth are often estimated, and patients can have multiple clinic identifiers that are not common within or across facilities [7]. These countries often do not have a single national identifier for all individuals, with national IDs given to those above a particular age and often not given to foreigners residing in the country [8]. Another challenge of using deterministic algorithms for patient matching is that they lack scalability, requiring expensive customization and business rule revisions as databases grow [5].

Probabilistic algorithms do not perform very well in many low-resource health care settings. In our institutional experience, the evaluation of various four-string manipulation strategies to improve the performance of probabilistic models, based on Kenyan names, revealed a suboptimal specificity and a positive predictive value of less than 50% [9]. While probabilistic algorithms are superior to deterministic algorithms, not all probabilistic algorithms applied to the same set of circumstances yield results with the same degree of accuracy [5]. This is because probabilistic algorithms typically sample the dataset and do not scan all possible values, thus matching functions become more complex and time consuming, increasing the number of false positive matches [10]. Current statistical matching models cause many challenges to unique patient identification in health care settings, and often tend to be difficult and expensive to implement. Thus, there is a critical need to evaluate relatively cheap, feasible, and effective solutions to tackle the patient matching problem in all health care settings, including LMICs and industrialized nations.

Biometric approaches offer a potential solution to the challenges of current patient matching algorithms. The basic principle of biometric authentication is that everyone is unique

and can be identified by his/her intrinsic physical or behavioral traits [11]. Among the available biometric technologies, fingerprint technology offers a potentially promising solution; fingerprint scanners are readily available, the technology is relatively, easy to use, has minimal database memory requirements, and there exist several demonstrated instances of their large-scale use in other sectors, such as banking and immigration departments.

Biometric technology is not without its challenges, among them being the need to invest in specialized technology and equipment required to capture many of the needed biometric measurements. Due to the poor performance of statistical patient matching models within our LMICs setting, we explored the potential of using fingerprinting biometric technology as additional metadata for patient matching. A fingerprint consists of a pattern of ridges and valleys in the surface of the fingertips and forming during early fetal months [12]. Apart from the algorithm, the performance of this technology is affected by factors such as image quality, composition of target user population, resolution, template size, and scanner type or model.

To determine the right device and approach for implementing biometrics, we conducted a systematic assessment of the technical performance of various fingerprinting devices available on the market that could be used within most care settings. The goal was to explore the selection of the right biometric device for countries aiming to develop systems for unique patient identification.

Methods

The envisioned workflow of the biometric technology at our institution, as in most clinical settings, will be as follows. When the patient first arrives at the facility, they will be asked to scan their left index finger through the fingerprint biometric scanner. Based on the scan, the fingerprint image will be converted to a template locally and this template will be sent over the network to the fingerprint matching service. The matching service will match the template against a database of existing fingerprints and will return NULL, if there is no match found or will return the patient id if a match is found. When NULL is returned, the patient registration page is opened and a new fingerprint image is used to create a fingerprint template for future matching. A matched fingerprint returns a patient id, allowing the patient's record to be directly accessed during the clinical encounter.

The current evaluation focused on identifying the technology that provides the best fingerprint match. To evaluate the technical performance of various scanners available in the market, we leveraged an application that worked with over 180 fingerprinting devices. This application, the Neurotech Verifinger Software Development Kit (SDK) 6.0, provides a wrapper of common Application Programming Interface (API) that makes it easy for use with different devices [13]. The goal of using this SDK was to showcase a methodology that other implementers could use when evaluating various scanners for their setting.

Fingerprint templates are mathematical representations of the most useful points of interest (minutiae) in fingerprint images [14]. Using fingerprint images of different fingers (not just the index finger), downloaded from various sources, including NIST SD4 [15], NIST SD9 [15], FVC2002 [16], FVC2004 [17], as well as randomly replicated of images, a large dataset of 50,000 images was generated. Using the SDK template API, we generated templates for these 50,000 images in the default format. This large number of images was adequate to stress test performance, as most care settings would have less

than this number of patients. Although the SDK supports ISO/IEC 19794-2:2005, ANSI/INCITS 378-2004, and ANSI/NIST-ITL 1-2007 standards, we chose to use the default one to avoid any bias between fingerprint readers. The test template dataset used one fingerprint image per template.

For generating the evaluation template, we used two images of the left index finger and then compared the performance and accuracy against the test dataset. There is good evidence that the index finger is used the least and hence has the least normal wear and tear, making it well-suited for fingerprint scanning [18]. Since we used the SDK to translate the minutiae to a template, all templates for both the test dataset and evaluation template used the same feature set and algorithm.

Each fingerprint scanner was used in the same surrounding lighting area to ensure the quality of the image capture was not affected. The devices were connected to an Ubuntu 16.04 x64 platform and tested using the C++ application. We used the standalone application instead of the web application that is based on an ActiveX component or Java applet, because both of these technologies have been deprecated by the browser manufacturers. The C++ application was also more robust in communication with the fingerprint scanners and responded more quickly to images from the fingerprint scanner. Live finger detection was also supported by the Futronic scanners only using the C++ application.

Eight Fingerprint devices (U.are.U 4500, U.are.U 4500 UID Edition, U.are.U 5160, Futronic/FS80, Futronic FS88H, Hamster Plus (HSDU03PTM), Biomini, and UPEK Eikon) were selected based on current community member usage and what was affordable and easily available for purchase. Key dimensions for the devices that were analyzed included: resolution, scan capture area, performance, template size, template format, compatible operating system, and supported standards.

Resolution

The number of pixels per inch (dpi) describing the acquired images. High scanner resolution allows for extraction of finer details from a fingerprint image, making it very important when identifying infants and elderly patients. A 500 dpi resolution is required by FBI-compliant systems [19].

Scan Capture Area

Determines the size of the fingerprint portion which can be acquired by the scanner. This parameter usually lies in the range 1.0"x1.0" square inches of some professional models to about 0.38"x0.38" of some low profile models. It is worth noting that the captured portion of the latter is about 7 times smaller than the former. A wide sensing area is important because the size of an average fingerprint is about 0.5"x0.7" (smaller for children and females and larger for adult males) and therefore the acquisition of a fingerprint with a sensing area smaller than 0.5"x0.7" produces a partial fingerprint [20].

Performance Matching

The speed and accuracy of identification and other derivatives that arise from accuracy. In our evaluation, performance was determined by the speed in milliseconds of correctly identifying a person from the test dataset of 50,000 fingerprint images. Matching speed impacts patient workflow and should be minimized to reduce patient waiting time during registration.

Template Size

Describes a stored file in a fingerprint scanning system and is normally stored as binary file. When a fingerprint is entered into the system, only a "template" of the fingerprint is stored, rather than the fingerprint image. A fingerprint template is

smaller than the actual fingerprint image and using the template instead of an image reduces processing time. Search speed is said to be directly related to template size; the smaller the template, the faster the search speed [21].

Template Format (Gray Scale Levels)

Of a fingerprint sensor is the number of gray shades produced for every pixel. 256 levels of gray is the standard format supported by most available fingerprint sensors today and results in using one-byte per pixel [22].

Compatible Operating System

Is highly dependant on the manufacturer of the scanner. Most scanners support Windows, Linux, iOS, and Android.

Supported Standards

This is similar to operating system compatibility and is largely determined by the manufacturer.

Results

Among the seven basic criteria for biometric security systems, performance or accuracy is a prerequisite (Table 1) [23]. The average performance time for the tested scanners was 1984.3 milliseconds, with wide variability across the different scanners. The best performance matching was achieved with the U.are.U 4500 UID edition, with a matching time of 600 milliseconds in a ratio of 1:50000 templates. The scanner resolution was 500 dpi, within the recommended range. The U.are.U 4500 scanner had a resolution higher than the U.are.U 4500 UID edition (512 dpi). This scanner will definitely produce a finer and more detailed extract from a fingerprint image, however its performance matching was 4-times higher than that of U.are.U 4500 UID edition. As performance is the most critical dimension for analysis, a scanner with the fastest performance matching speed and minimum resolution requirement (i.e., 500 dpi) is preferred.

We observed that search speed when using fingerprint biometric devices was directly related to template size (i.e., the smaller the template, the faster the search speed) [19]. There appeared to be a trade-off among the template size, scan capture area, and performance matching. This is because the fastest scanner (U.are.U 4500 UID edition) did not have the smallest template (72 mm x 39 mm x 21.7). This particular device also did not have a large scan capture area (12.8 mm x 16.5 mm). While this was desirable, manufacturing large and pure silicon chips is difficult and rather expensive; therefore, the scanners currently available on the market are categorized by a small area scan capture area.

On average, it would cost about \$107.60 to acquire a scanner (the cost of different fingerprint scanners has been provided in Table 1), though cost is crucial while acquiring these gadgets, we can not compromise performance for cost. That is why an inexpensive scanner with a good resolution, but poor performance matching, may not be a better choice.

Discussion

Biometrics offers an alternative method to collect additional metadata for statistical patient matching models, with good performance characteristics. When determining what scanner to buy, organizations should consider scanners with short turnaround time. We demonstrate the need to critically evaluate multiple dimensions of available scanners prior to purchase. Further, issues around what SDK to use, and how to integrate scanning within the clinical workflow is important and could be a limiting factor for scanner adoption.

Our results demonstrate that the scan area has an impact on performance speed. Optimizing scan area can be improved by training and providing ambient conditions to capture fingerprint images. The technique of capturing the fingerprint may necessitate a large scan area (e.g., in cases where you want to simultaneously capture multiple fingerprints). When speed is optimized, template size can be tweaked to reduce memory and storage requirements. This is particularly useful for mobile health where there are database capacity and infrastructure limitations.

Privacy and security implications surrounding the use of biometrics is extremely important, but was beyond the scope of the current study. Given that a number of large-scale collections of fingerprints is already underway (e.g., passports and social security in both high-income and low-income countries), debates are largely dependent on the reason for capturing biometric information. Identity theft, no means to revoke leaked biometrics, health equity, denying services to individuals, and many other ethical issues are some of the factors affecting the success of biometric implementations.

There are limitations to the current study. Only a few devices were evaluated; we did not separate groups of patients (e.g., neonates and elderly), who may skew the analysis and variation of performance between these groups; and we did not evaluate performance of various fingers (e.g., index versus middle finger for a specific patient). The evaluation was also conducted using a single test set; findings would likely be different based on the population being evaluated.

Numerous costs impact the implementation of biometric systems and vary by the size of the organization, choice of system adopted (open source or commercial), and the personnel to manage the infrastructure. Pre-scan enhancer pads, which cost roughly \$47, may also be needed. Fingerprint SDK, Neurotechnology SDK costs about \$422 for VeriFinger 9.0 Standard SDK. Commercial SDKs are usually restricted to specific scanners however, open source SDKs are also available and are not scanner-specific. Note, that even commercial SDKs require personnel for set-up and maintenance services. Personnel costs (i.e., both training and hire) are highly dependent on the complexity of the system being deployed. Finally, on the back end, most biometric applications are hosted on a server and accessed via client machines; machines should be purchased with a strong firewall to ensure protection from hackers.

Conclusion

Biometric fingerprint scanners offer one potential technology for improving patient matching and unique patient identification in diverse health care settings. Special attention is needed when selecting these technologies to ensure good performance at a reasonable cost. Additionally, it is important that the technology can be implemented within existing workflows and with consideration to existing patient care setting infrastructure constraints. This is especially challenging in clinical settings with complex health care workflows. While this paper is geared towards LMICs, these results and findings are equally applicable to care settings in industrialized nations.

Acknowledgements

This work was supported in part by the NORHED program (Norad: Project #QZA-0484, HI-TRAIN Project), Grand Challenges Canada Stars in Global Health - Round 5 Phase I (S5 0416-01). The content is solely the responsibility of the

authors and does not necessarily represent the official views of the Norwegian Agency for Development Cooperation.

References

- [1] Safety and Quality Improvement Guide Standard 5: Patient Identification and Procedure Matching (October 2012), in, Australian Commission on Safety and Quality in Health Care, Sydney, 2012.
- [2] B.H. Just, D. Marc, M. Munns, and R. Sandefer, Why Patient Matching Is a Challenge: Research on Master Patient Index (MPI) Data Discrepancies in Key Identifying Fields, *Perspectives in Health Information Management* **13** (2016).
- [3] F. Verbeke, S.V. Bastelaere, and M. Nyssen, Patient Identification And Hospital Information Management Systems In Sub-Saharan Africa: A Prospective Study In Rwanda And Burundi., *Rwanda Health Communication Center - Rwanda Biomedical Center (RHCC - RBC)* **69** (2013), 7-12.
- [4] L.M. Gliklich and R.E. Dreyer, Registries for Evaluating Patient Outcomes: A User's Guide [Internet], in, 2014.
- [5] S. Schumacher (2007) *Probabilistic versus deterministic data matching: making an accurate decision, DM Direct*.
- [6] *Deterministic and Probabilistic Data Matching (Master Index Match Engine Reference)*. Retrieved from https://docs.oracle.com/cd/E19182-01/821-0919/ref_sme-deter-probl_c/index.html.
- [7] Oracle (2010) *Deterministic and Probabilistic Data Matching (Master Index Match Engine Reference)*. Retrieved from https://docs.oracle.com/cd/E19182-01/821-0919/ref_sme-deter-probl_c/index.html.
- [8] theBeehive (2001) *How to get a National Identity Card | The Beehive - Kenya*. Retrieved from <http://kenya.thebeehive.org/en/content/640/1757>.
- [9] P. Sonak, W. Martin, and G.J. Wawira, Improving registration to improve retention and linkage of patients for rural healthcare delivery in Western Kenya, in, 2014.
- [10] K. Perspectives (2016) *Deterministic versus Probabilistic Matching in Big Data - Knowledge Perspectives*. Retrieved from <https://blog.knowledgent.com/deterministic-versus-probabilistic-matching-big-data/>.
- [11] M. Cobb. *What is biometrics?* Retrieved from <http://searchsecurity.techtarget.com/definition/biometrics>.
- [12] H. Jhaveri, H. Jhaveri, and D. Sanghavi, Biometric Security System And Its Applications in Healthcare, *International Journal of Technical Research and Applications* **2** (2014), 15-20.
- [13] VeriFinger fingerprint recognition technology, algorithm and SDK for PC, smartphones and Web. Retrieved from <http://www.neurotechnology.com/verifinger.html>.
- [14] X. Jiang and Y. Wei-Yun, Fingerprint minutiae matching based on the local and global structures, in: *International Conference on Pattern Recognition*, IEEE, 2000, pp. 1038-1041.
- [15] NIST. *NIST Special Database 4 | NIST*. Retrieved from <https://www.nist.gov/srd/nist-special-database-4>.
- [16] FVC2002 - Second International Fingerprint Verification Competition. Retrieved from <http://bias.csr.unibo.it/fvc2002/download.asp>.
- [17] FVC2004 - Third International Fingerprint Verification Competition. Retrieved from <http://bias.csr.unibo.it/fvc2004/download.asp>.
- [18] Desert Hand Therapy. 18 Amazing Facts About Human Hands | Desert Hand Therapy, in, 2016.
- [19] M. Sandstrom, C. Fredrik, and V. Fak, Liveness Detection in Fingerprint Recognition Systems, (2004).
- [20] Biometrika srl (2015) *Biometrika - Basics of fingerprint recognition technology and biometric systems*. Retrieved from http://www.biometrika.it/eng/wp_scfing.html.
- [21] F.W.S. Bhd (2009) FingerTec. *Fingerprint Technology White Paper*. Retrieved from <http://www.fingertec.com/academia/download/whitepaper-01.pdf>.
- [22] Next Biometrics Group Asa, *Resolution & Gray Scale Levels - NEXT Biometrics*. Retrieved from http://nextbiometrics.com/security/resolution_gray_scale_levels/.
- [23] R. Jain, Abstract : Keyword : Table of Contents : 2-Application Fields for Biometrics Technology, (2004), 1-10.

Address for correspondence

Noah J. Kasiiti

College of Health Sciences, Institute of Biomedical Informatics, Moi University, Eldoret, Kenya, P.O Box 3900-30100

Email:kasiitinoah@gmail.com,

Tel1: +25479016388, Tel2: +256753764102

Table 1: Performance analysis of various fingerprint scanners

Brands	Resolution (dpi)	Template Format (bit grayscale (256 gray levels))	Template Size (mm)	Scan Capture Area (mm)	Supported Standards	Performance (ms, Matching 1:50,000)	Operating System Compatibility	Cost (USD)
U.are.U 4500	512	8	65 x 36 x 15.6	14.6 x 18.1	FCC Class B, CE, ICES, BSMI, MIC, USB, WHQL	2420	Microsoft Windows, Linux	\$75.00
U.are.U 4500 UID Edition	500	8	72 x 39 x 21.7	12.8 x 16.5	FIPS 201 PIV, STQC	600	Microsoft Windows, Linux	\$60-70
U.are.U 5160	500	8	72 x 39.5 x 21.7	15 x 18	FIPS 201 PIV, RoHS, WEEE UL, USB, WHQL	1388	Windows Linux Android	\$183.81
Futronic' FS80	500	8	45 x 63 x 26	16 x 24	USB 2.0 compatible interface, plug and play device	1450	Windows, Linux, MAC OS , Android	\$65.99
Futronic FS88H	500	8	59 x 60 x 20	16.3 x 24.4	FIPS 201/PIV, FAP20, Microsoft WHQL, FCC and CE, RoHS	1280	Windows, Linux, MAC OS , Android	\$233.92
Hamster Plus (HSDU03P™)	500	8	53 x 73 x 84	13.2 x 15.2	FCC, CE, KCC, RoHS	900	Windows, Windows Server, Android, Java, Linux	\$76.99
Biomini	500	8	66 x 90 x 58	16 x 18	CE, FCC, KC, WHQL	3800	Windows, Linux, Android	\$115.00
UPEK Eikon	508	8	84 x 34 x 14	25 x 10	• ISO/IEC 19794-2 (minutiae) and ISO/IEC 19794-4 (image) • ANSI INCITS 378 (minutiae) and ANSI INCITS 381 (image), • WSQ 3.1	2952	Windows, Linux, MAC OS , Android	\$ 39.95

Development and Evaluation of a Blood Glucose Management System for Reducing the Delay in Measurement

Teramoto Kei^a, Okura Tsuyoshi^b, Koshitani Rie^c, Takahiro Murata^a, Yorifuji Yuko^c,
Hotta Erika^d, Endo Yusuke^e, Ueki Masaru^e, Kondoh Hiroshi^a

^a Division of medical informatics, Tottori University Hospital, Yonago City, Japan

^b Division of Molecular Medicine and Therapeutics, Department of Multidisciplinary Internal Medicine, Tottori University Faculty of Medicine Yonago City, Japan

^c Division of Nursing, Tottori University Hospital Yonago City, Japan

^d eBase Solutions Laboratory, Inc. Yonago City, Japan

^e Center for Promoting Next-Generation Highly Advanced Medicine, Tottori University Hospital Yonago City, Japan

Abstract

The purpose of this study was to develop and evaluate the blood glucose management (BGM) system, which supports the scheduling of blood glucose measurements for medical staff. This system enables medical staff to continually confirm the instructions of the physicians by measuring blood glucose levels using a smart device (iPod). This paper describes the difference in the delay of the measurement between the BGM system and the non-BGM system. For the iPod device in the BGM system, the desktop device, and the laptop device, the frequencies of a delay under 30 minutes were 94%, 67%, and 80%, respectively (Ryan-test, $P < 0.01$), and that over 30 minutes were 6%, 33%, and 20%, respectively (Ryan-test, $P < 0.01$). We concluded that the BGM system reduced the delay in blood glucose measurements and led to an optimization of the time of measurement.

Keywords:

Electronic Health Records; Hospital Information Systems; Blood Glucose

Introduction

It is well-known that diabetes is one of the most common metabolic disorders. In 2014, the International Diabetes Federation (IDF) estimated that 387 million people had diabetes in IDF regions, which consists of 194 countries [1-9]. In Japan, there were 7.2 million patients with diabetes in 2015. The Japanese Ministry of Health, Labor and Welfare reported that 16% of all inpatients had diabetes in 2014.

Blood glucose control is conducted to keep levels of blood glucose in the normal range for a patient with diabetes. Establishing a normal blood glucose level for inpatients with diabetes is important to prevent complications and infections after surgery [10-24]. For blood glucose control, medical staff have to frequently measure blood glucose levels and register this data in the electronic medical record (EMR) system [25-28]. On the screen of the EMR system, medical staff confirm the instructions ordered by physicians, such as the insulin dose. The time to respond to changes in blood glucose levels is dependent on the difference between the actual recorded time and the scheduled time of measurements of blood glucose levels.

When medical staff register blood glucose levels to the EMR system manually, delay in recording, transcription errors and

patient misidentification occur [29-30]. In recent years, to prevent these errors, medical vendors have developed a blood glucose tester that transfers the results of the blood glucose test to the EMR system through a wireless network [31]. Although the tester prevents transcription errors of blood glucose levels, it does not prevent the delay in measuring blood glucose levels. The primary reason for the delay is that understanding various measurement orders from physicians is difficult for medical staff. In this study, to reduce the delay in recording blood glucose levels, the author developed the blood glucose management (BGM) system that enables medical staff to understand the schedules of the blood glucose control ordered by physicians.

The purpose of this study was to develop and evaluate the BGM system that supports the scheduling of blood glucose measurements for medical staff.

Study setting

DAISEN is an EMR system comprising various departmental healthcare systems for all outpatients and inpatients in the Tottori university hospital. As an EMR system, DAISEN records data of blood glucose measurements linked with the instructions ordered by physicians. Use of clinical data from DAISEN for this study was approved by the institutional review board of the Tottori university hospital.

The data for this study were extracted from the records in the EMR system, which consisted of blood glucose levels of inpatients with type 2 diabetes who were admitted to the hospital between Jan 1, 2014 and Oct 1, 2016. The number of selected patients before installation of the BGM system was 3,285 and the number of selected patients after installation of the BGM system was 2,847. To extract the delay in measuring blood glucose levels, we obtained the time that physicians requested the measurement using the EMR system and the time that the blood glucose level, measured by the blood glucose tester, was recorded by medical staff in the EMR system. The authors defined the delay in recording blood glucose levels as the difference between the requested and recorded time. We obtained 35,836 and 9,698 data records with desktop and laptop devices, respectively, as evaluation data before installation of the BGM. After installation of the BGM system, we obtained 93,639 data records with smart devices (iPod).

Evaluation of the BGM system

Analytical processes of evaluation for BGM system are as follows.

Delay time of recording blood glucose level to EMR system

To measure the delay in transferring blood glucose levels into the BGM system and in the non-BGM system, the authors defined the delay as the difference between the time that the blood glucose measurements were recorded by medical staff and the time that the measurement was requested by physicians.

The Occurrence rate of missed measurement or recording of blood glucose levels

To calculate the occurrence rate of non-existing blood glucose value in the EMR system, the author defined missed measurement or recording of blood glucose level as when medical staff did not measure blood glucose level within 60 minutes from the designated time by the physician.

Experimental devices

In this study, the authors developed the BGM system, which consisted of a blood glucose tester, an iPod, and a blood glucose data transfer adaptor (Table 1).

The blood glucose tester was used to measure the blood glucose level of inpatients with diabetes. The data transfer adaptor was used to transfer the data of the blood glucose levels and the execution time recorded by the blood glucose tester to an iPod.

Medical staff used an iPod to transfer the data to the EMR system. Functions of the application installed on the iPod to support the scheduling of blood glucose measurements are described in Table 1.

Table 1 – Experimental devices.

Device	Hardware/Operating System
Application	• iPod touch A1421/iOS 9.3
Blood glucose tester	• One Touch Verio pro plus
Data transfer adaptor	• ebase data transfer adapter P-model
EMR system	• IBM Power System E870/AIX(Clinical Information system)

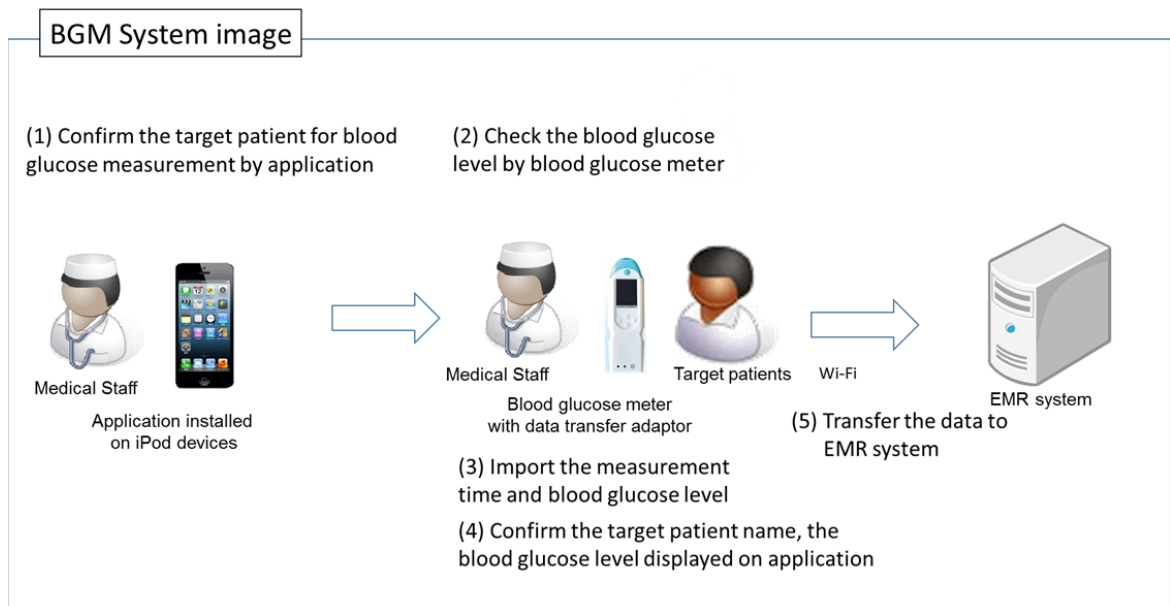


Figure 1 – Blood glucose measurement by the BGM system

With the use of the Objective-C programming language, the application was developed to display the list of selected patients linked with instructions of physicians and send messages to medical staff to alert them of any delay in recording blood glucose levels. The list of selected patients was extracted from the EMR data using Structured Query Language (SQL). The alert made medical staff carrying an iPod aware of a 15 or 30 minutes' delay in the measurement using sounds or vibration.

Blood glucose measurement with the BGM system consists of five steps (Figure 1).

1. Medical staff confirm the selected patient who needs a blood glucose measurement by using an application on the iPod.

2. Medical staff check the blood glucose level of a target patient using a blood glucose tester with the data transfer adaptor.

3. Medical staff import the measurement time and the blood glucose level from the data transfer adaptor into the iPod to display the data on the screen.

4. Medical staff confirm the selected patient's name, the blood glucose level, and the measurement time displayed on the screen before registration to the EMR system.

5. The application transfers the data to the EMR system through a wireless network.

In the non-BGM system, medical staff manually recorded blood glucose levels to the EMR using a desktop or laptop

device by transferring the measurement values displayed on the blood glucose tester. All desktop devices were located in a nursing station. Laptop devices were placed in a computer cart designed for laptops, which can be carried to the inpatients' bedsides.

Statistical analysis

The delay of 15 and 30 minutes of blood glucose measurements for the BGM system and non-BGM system was calculated to compare the occurrence rates using the Ryan test. P values <0.05 were considered statistically significant. The R software program (version 3.1.2) was used to perform the statistical analysis.

Results

Results of delay time to recording the value to EMR system from measuring blood glucose level

The installation of the BGM system resulted in a significant reduction in the delay of recording blood glucose levels by medical staff. Table 2 show the occurrence rate of a 15 minute delay in recording blood glucose levels when using an iPod device, desktop device and laptop device. When using an iPod in the BGM system, a desktop device, and a laptop device, frequencies of a delay under 15 min were 77%, 37%, and 53%, respectively (Ryan test, $P < 0.01$); frequencies of a delay under 30 min were 94%, 67%, and 80%, respectively (Ryan test, $P < 0.01$); and frequencies of a delay over 30 minutes were 6%, 33%, and 20%, respectively (Ryan test, $P < 0.01$).

Table 2 - The occurrence rate of delay time for each device.

Devices	Median (h:m:s) * ₁	Average * ₂	SD * ₃	Total* ₄	Occurrence frequency of a delay of each 15 min			Ryan test (< 15 min vs. > 15 min)	Ryan test (< 30 min vs. > 30 min)
					< 15 (%) * ₅	15-30 (%)	> 30 (%)		
iPod (BGM)	0:04:37	0:09:07	0:11:14	93764	71702 (76)	16003 (17)	6059 (6)	vs. Desktop $p < 0.01$	vs. Desktop $p < 0.01$
Desktop PCs (non-BGM)	0:21:10	0:23:15	0:16:24	9709	3610 (37)	2961 (30)	3138 (32)	vs. Laptop $p < 0.01$	vs. Laptop $p < 0.01$
Laptop PCs (non-BGM)	0:17:56	0:18:38	0:15:23	35904	18943 (53)	9598 (27)	7363 (21)	vs. iPod $p < 0.01$	vs. iPod $p < 0.01$

*1 Median of the difference between the time that the blood glucose measurements were performed by medical staff and the time that the measurement was requested by physicians.

*2 Average of the difference between the time that the blood glucose measurements were performed by medical staff and the time that the measurement was requested by physicians.

*3 Standard deviation of the difference between the time that the blood glucose measurements were performed by medical staff and the time that the measurement was requested by physicians.

*4 The total number of blood glucose measurements.

*5 The occurrence rate was defined as the occurrence of a delay of 15 minutes divided by the total occurrence rate recorded for blood glucose levels.

The occurrence rate of missed measurements or recordings

We calculated the occurrence rates of missed measurements or recordings of a blood glucose level for each device environment. As a result, for desktop devices, laptop devices, and iPod devices the occurrence rates are 23%, 16% and 5% respectively. The results for iPod were significantly lower than those for desktop and laptop devices. ($P < 0.01$, Ryan test)

Discussion

In this study, we showed that our BGM system significantly reduced the delay in recording blood glucose levels and the occurrence rate of missed measurements or recordings compared with non-BGM systems. The results are expected to contribute to accurate blood glucose control. A large number of inpatients after surgery are regulated by blood glucose control with insulin injections. Medication errors with insulin have the potential to result in serious harm, including death [32-37]. When blood glucose control was combined with insulin, measurement or record time of glucose levels was set based on the time of a meal. The time of a meal of inpatients was influenced by the measurement of blood glucose levels because the measurement had to be done before or after the meal. The BGM system notifies medical staff of the instructions by physicians such as the insulin dose.

Consequently, there is a possibility of improving medical safety of patients with diabetes by reducing the delay of measurement using the BGM system.

Effectiveness of an iPod in availability and accessibility in the medical field has been reported in recent studies [38]. This study focused on the effectiveness of its use for the measurement of blood glucose levels by medical staff. The results showed that the medical staff with iPod devices carried out the measurement more accurately and recorded the blood glucose levels quicker than the medical staff with other devices. More specifically, the results were interesting in that they showed superiority of an iPod compared to laptop PCs though both devices can easily be carried to the bedside of selected patients.

To prevent a patient misidentification accident, medical records of only one patient can be displayed in a screen of the EMR system. On the other hand, all selected patients who needed blood glucose measurements were displayed in the order they were seen by the physician on the screen of the iPod with the BGM system to allow medical staff to effectively take care of many inpatients in a ward. It is possible that the display of selected patients as described above led to the superiority of an iPod compared to laptop PCs.

Compared with the occurrence rate of missed measurements or recordings for each device, we found that the BGM system contributed to decreasing missed measurements of blood

glucose level. However, the results did not sufficiently clarify the effect on medical safety for real inpatient care. For evaluation of medical safety in clinical settings, further investigations based on other resources such as incident reports by medical staff would be needed.

The application on an iPod enabled medical staff to usually understand the situation of the blood glucose measurement. Thus, it is desirable that all medical staff who measure blood glucose levels always carry an iPod. Although the number of hospitals adopting smart devices such as an iPod for medical use has increased in the recent years, its usage is still considered to be insufficient. Thus, the cost of the BGM system including smart devices remains an issue.

In the future, from the practical perspective, implementation costs need to be discussed to use the BGM system in other medical facilities.

Conclusion

We concluded that the BGM system reduced the delay of measurement and resulted in an optimization of the time of measurement. With the increased use of smart devices such as the iPod in the medical field, the system will likely be implemented to the EMR system in many hospitals.

We conclude that for the iPod device, desktop device, and laptop device, the frequencies of a delay under 30 minutes were 94%, 67%, and 80%, respectively (Ryan test, $P < 0.01$), and over 30 minutes were 6%, 33%, and 20%, respectively (Ryan test, $P < 0.01$).

Transcription errors of the blood glucose level were found in the EMR system before installation of the BGM system. The BGM system can prevent transcription errors of recording blood glucose levels using a data transfer adapter.

References

- [1] A.F. Amos, D.J. McCarty, and P. Zimmet, The rising global burden of diabetes and its complications: estimates and projections to the year 2010, *Diabet Med* **14 Suppl 5** (1997), S1-S85.
- [2] J. da Rocha Fernandes, K. Ogurtsova, U. Linnenkamp, L. Guariguata, T. Seuring, P. Zhang, D. Cavan, and L.E. Makaroff, IDF Diabetes Atlas estimates of 2014 global health expenditures on diabetes, *Diabetes Res Clin Pract* **117** (2016), 48-54.
- [3] G. Danaei, M.M. Finucane, Y. Lu, G.M. Singh, M.J. Cowan, C.J. Paciorek, J.K. Lin, F. Farzadfar, Y.H. Khang, G.A. Stevens, M. Rao, M.K. Ali, L.M. Riley, C.A. Robinson, M. Ezzati, and G. Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating, National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants, *Lancet* **378** (2011), 31-40.
- [4] L. Guariguata, D.R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J.E. Shaw, Global estimates of diabetes prevalence for 2013 and projections for 2035, *Diabetes Res Clin Pract* **103** (2014), 137-149.
- [5] H. King, R.E. Aubert, and W.H. Herman, Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections, *Diabetes Care* **21** (1998), 1414-1431.
- [6] H. King and M. Rewers, Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. WHO Ad Hoc Diabetes Reporting Group, *Diabetes Care* **16** (1993), 157-177.
- [7] W. Rathmann and G. Gianni, Global prevalence of diabetes: estimates for the year 2000 and projections for 2030, *Diabetes Care* **27** (2004), 2568-2569; author reply 2569.
- [8] D.R. Whiting, L. Guariguata, C. Weil, and J. Shaw, IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030, *Diabetes Res Clin Pract* **94** (2011), 311-321.
- [9] A. Alrwisan, J. Ross, and D. Williams, Medication incidents reported to an online incident reporting system, *Eur J Clin Pharmacol* **67** (2011), 527-532.
- [10] R.E. Amori, A.G. Pittas, R.D. Siegel, S. Kumar, J.S. Chen, S. Karnam, S.H. Golden, and D.N. Salem, Inpatient medical errors involving glucose-lowering medications and their impact on patients: review of 2,598 incidents from a voluntary electronic error-reporting database, *Endocr Pract* **14** (2008), 535-542.
- [11] M. Garrouste-Orgeas, J.F. Timsit, A. Vesin, C. Schwebel, P. Amodo, J.Y. Lefrant, B. Souweine, A. Tabah, J. Charpentier, O. Gontier, F. Fieux, B. Mourvillier, G. Troche, J. Reignier, M.F. Dumay, E. Azoulay, B. Reignier, J. Carlet, L. Soufir, and O.S. Group, Selected medical errors in the intensive care unit: results of the IATROREF study: parts I and II, *Am J Respir Crit Care Med* **181** (2010), 134-142.
- [12] T. Lamont, D. Cousins, R. Hillson, A. Bischler, and M. Terblanche, Safer administration of insulin: summary of a safety report from the National Patient Safety Agency, *BMJ* **341** (2010), c5269.
- [13] J.P. Santell, R.W. Hicks, J. McMeekin, and D.D. Cousins, Medication errors: experience of the United States Pharmacopeia (USP) MEDMARX reporting system, *J Clin Pharmacol* **43** (2003), 760-767.
- [14] E.S. Schwenk, B. Mraovic, R.P. Maxwell, G.S. Kim, J.M. Ehrenfeld, and R.H. Epstein, Root causes of intraoperative hypoglycemia: a case series, *J Clin Anesth* **24** (2012), 625-630.
- [15] A.N. Thomas and U. Panchagnula, Medication-related patient safety incidents in critical care: a review of reports to the UK National Patient Safety Agency, *Anaesthesia* **63** (2008), 726-733.
- [16] A.G. Winterstein, R.C. Hatton, R. Gonzalez-Rothi, T.E. Johns, and R. Segal, Identifying clinically significant preventable adverse drug events through a hospital's database of adverse drug reaction reports, *Am J Health Syst Pharm* **59** (2002), 1742-1749.
- [17] L.a.W. Japan Ministry of Health, <http://www.mhlw.go.jp/toukei/saiken/hw/kanja/14/dl/toukei.pdf> (accessed 2016 Dec 16), (2014).
- [18] A. Abou-Saleh, M. Haq, and D. Barnes, Inpatient management of diabetes in adults: safety and good practice, *Br J Hosp Med (Lond)* **75** (2014), 258-263.
- [19] P. Arnold, D. Scheurer, A.W. Dake, A. Hedgpath, A. Hutto, C. Colquitt, and K.L. Hermayer, Hospital Guidelines for Diabetes Management and the Joint Commission-American Diabetes Association Inpatient Diabetes Certification, *Am J Med Sci* **351** (2016), 333-341.
- [20] Z.T. Bloomgarden, Inpatient diabetes control - Approaches to treatment, *Diabetes Care* **27** (2004), 2272-2277.
- [21] T.G. Breuer and J.J. Meier, Inpatient treatment of type 2 diabetes, *Dtsch Arztebl Int* **109** (2012), 466-474.
- [22] S. Brumm, K. Theisen, and M. Falciglia, Diabetes Transition Care From an Inpatient to Outpatient Setting in a Veteran Population: Quality Improvement Pilot Study, *Diabetes Educ* **42** (2016), 346-353.
- [23] D.L. Carpenter, S.R. Gregg, K. Xu, T.G. Buchman, and C.M. Coopersmith, Prevalence and Impact of Unknown Diabetes in the ICU, *Crit Care Med* **43** (2015), e541-550.
- [24] J.J. Hsu, J.I. Wang, A. Lee, D.Y. Li, C.H. Chen, S. Huang, A. Liu, B.K. Yoon, S.K. Kim, and T.J. Tsai, Automated control of blood glucose in the OR and surgical ICU, *Conf Proc IEEE Eng Med Biol Soc* **2009** (2009), 1286-1289.
- [25] N.B. Khazai and O. Hamdy, Inpatient Diabetes Management in the Twenty-First Century, *Endocrinol Metab Clin North Am* **45** (2016), 875-894.
- [26] M. Michalia, M. Kompoti, A. Koutsikou, A. Paridou, P. Giannopoulou, E. Trikkia-Graphakos, and P. Clouva-Molyvdas, Diabetes mellitus is an independent risk factor for ICU-acquired bloodstream infections, *Intensive Care Med* **35** (2009), 448-454.
- [27] S.K. Min, J.H. Shin, M.Y. Chang, H.J. Min, K.S. Kim, S.Y. Lee, H.S. Yang, Y.H. Hong, and S.K. Mun, Impact of control of blood glucose level during treatment of sudden deafness in diabetics: relationship with prognosis, *Eur Arch Otorhinolaryngol* (2016).
- [28] R.J. Rushakoff, Inpatient Diabetes Management, in: *Endotext*, L.J. De Groot, G. Chrousos, K. Dungan, K.R. Feingold, A. Grossman, J.M. Hershman, C. Koch, M. Korbonits, R. McLachlan, M. New, J. Purnell, R. Rebar, F. Singer, and A. Vinik, eds., South Dartmouth (MA), 2000.
- [29] C. Smallwood, D. Lamarche, and A. Chevrier, Examining Factors That Impact Inpatient Management of Diabetes and the Role of Insulin Pen Devices, *Can J Diabetes* (2016).
- [30] G.E. Umpierrez, R. Gianchandani, D. Smiley, S. Jacobs, D.H. Wesorick, C. Newton, F. Farrokhi, L. Peng, D. Reyes, S. Lathkar-Pradhan, and F. Pasquel, Safety and efficacy of sitagliptin therapy for the inpatient management of general medicine and surgery patients with type 2 diabetes: a pilot, randomized, controlled study, *Diabetes Care* **36** (2013), 3430-3435.
- [31] D.J. Wexler and E. Cagliero, Inpatient diabetes management in non-ICU settings: evidence and strategies, *Curr Diabetes Rev* **3** (2007), 239-243.
- [32] D.J. Zou, [Operation treatment and medicine treatment of type 2 diabetes mellitus combined with obesity], *Zhonghua Wei Chang Wai Ke Za Zhi* **12** (2009), 543-544.
- [33] M. Greiver, K. Keshavjee, K. Martin, and B. Aliarzhadeh, Who are your patients with diabetes?: EMR case definitions in the Canadian primary care setting, *Can Fam Physician* **58** (2012), 804, e421-802.
- [34] J.E. Hernandez-Avila, A. Lara, E. Morales-Carmona, E.G. Espinoza, P. Anaya, and L.S. Palacio-Mejia, Extracting And Using Data From Electronic Medical Records (Emr) To Monitor Quality Of Care And Prescription Patterns For Diabetes Prevention And Control In Outpatient

- Clinics Of Low And Mid Resources Countries: The Case Of Colima, Mexico, *Value Health* **18** (2015), A811.
- [35] M.T. Korytkowski, M. Brooks, M. Lombardero, D. DeAlmeida, J. Kanter, V. Magaji, T. Orchard, and L. Siminerio, Use of an Electronic Medical Record (EMR) to Identify Glycemic Intensification Strategies in Type 2 Diabetes, *J Diabetes Sci Technol* **9** (2015), 593-601.
- [36] D.G. Morrow, T. Conner-Garcia, J.F. Graumlich, M.S. Wolf, S. McKeever, A. Madison, K. Davis, E.A. Wilson, V. Liao, C.L. Chin, and D. Kaiser, An EMR-based tool to support collaborative planning for medication use among adults with diabetes: design of a multi-site randomized control trial, *Contemp Clin Trials* **33** (2012), 1023-1032.
- [37] H.-L.R. Ltd,
http://www.rochecanada.com/en/products/diagnostics/doctors-nurses/decentralized_solutions/blood-glucose-testing.html(accessed)
- [38] Greysen, S. R., et al. . "Tablet computers for hospitalized patients: a pilot study to improve inpatient engagement." *J Hosp Med* **9**(6): 396-399.

Address for correspondence

Kei Teramoto

Division of medical informatics, Tottori University Hospital, 36-1 Nishi-CHO, Yonago City, 683-8504 Japan

Email: kei-tera@umin.net

Tel: +81-859-38-7128, Fax: +81-859-38-6899

Evaluation of a Decision Support System for Laboratory Service Patients

Georgy Kopanitsa^{a,b}

^a*Institute Cybernetic Center, Tomsk Polytechnic University, Tomsk, Russia*

^bTomsk State University for Architecture and Building, Tomsk, Russia

Abstract

In Russia, many patients address laboratory services directly without a doctor's referral. This causes the problem of interpretation of laboratory test results by the patients who don't have a proper medical background. We have a situation when there is no healthcare professional between a patient and the test results. This problem can be solved if a laboratory service provides not only the results of the tests but also their interpretations. The system was implemented in the Helix laboratory service in Saint-Petersburg, Russia. At the moment it generates about 3500 reports a day. This study presents the evaluation of the system, which has shown a good performance both on correctness of generated reports and user acceptance.

Keywords:

Decision Support; Clinical Laboratory Information Systems;
Telemedicine

Introduction

A need of a decision support system that would help patients to read and understand laboratory test results comes from the fact that many patients address laboratory services directly without a doctor's referral [1; 2]. This causes the problem of interpretation of laboratory test results by the patients who don't have a proper medical background and leads to a situation when there is no healthcare professional between a patient and the test results [3]. This problem can be solved if a laboratory service provides not only the results of the tests but also their interpretations. Automated decision support systems that have proved their efficiency for doctors can be a good solution for this problem [4]. The experience of development and implementation of decision support systems for doctors shows the efficiency of such solutions for the doctors, however, developers face problems when it comes to the decision support for patients. They require different approach in data presentation and interpretation [5-8]. The DoctorEase system was implemented in the Helix laboratory service in Saint-Petersburg, Russia. At the moment it generates about 3500 reports a day. The example of a report is presented in the figure 1.

The goal of this paper is to present the evaluation of correctness and user acceptance of a decision support system for the patients of a laboratory service.



Лаборатория служба Кликс
 Телефон: 8(800) 700-03-03
 Информация в интернете: www.helix.ru

Лицензия ЛО-78-01-015-2015
 КСД-000-2008 ГОСТ Р ISO 15189:2009
 Код в реестре выданных сертификатов качества (СЗД) 0401
 Код в реестре выданных сертификатов качества (СЗСХ) 1671

Физико-химические показатели общего анализа мочи:

Интерпретация кислотности (реакции) мочи, удельного веса мочи в связи с их высокой вариабельностью и зависимостью от многих факторов, в том числе объема потребляемой жидкости, качественного состава потребляемых продуктов, по данным только одного общего анализа мочи не проводится.

В общем анализе мочи по всем физико-химическим показателям не выявлено никаких значимых отклонений от нормы.

Микроскопическое исследование осадка мочи:

При проведении микроскопического осадка мочи в Вашем общем анализе мочи выявлены оксалаты (соли щавелевой кислоты, образующиеся в почках).

Наличие оксалатов в моче может быть вызвано нарушением минерального состава мочи или большим количеством продуктов, богатых щавелевой кислотой, в Вашем рационе (щавель, шпинат, бобовых, ячменя, каша, шоколад, крепкого чая, натурального кофе и др.).

Поэтому однократное выявление оксалатов в общем анализе мочи не может однозначно считаться признаком нарушения минерального состава мочи. Для подтверждения наличия данного нарушения необходимо как минимум 3-4 кратное последовательное определение оксалатов в общем анализе мочи.

Вам необходимо ограничить потребление продуктов, богатых щавелевой кислотой, и сдать повторный анализ мочи через 3-7 дней жизни.

Повышенное содержание оксалатов в моче может стать предположительным фактором для развития мочекаменной болезни.

Поэтому, если в двух подряд анализах мочи у Вас выявляются оксалаты, следует посетить уролога или нефролога, который выявит причину данного состояния и решит вопрос ее коррекции.

Врач:

Заведующий лабораторией:

Дата: 05.02.2015



А. Яковенко

И. Скибо

Figure 1 – Report example

Methods

Correctness of the decision support

To evaluate the correctness of the generated reports a validation of a sample of 256 reports was sent to two independent laboratory service physicians for independent expert review. The results of the review was used to calculate precision ((All terms – Mistakes)/All terms), recall (ratio of true positives to (true positives + false negatives)), and F-measure ($2 \cdot (\text{recall} \cdot \text{precision}) / (\text{recall} + \text{precision})$) [9]. All the disagreements between reviewers were settled by consensus. Cohen's kappa has been calculated to rate the disagreement between experts [10].

User acceptance

To evaluate the user acceptance of the decision support system we have applied a Wilson and Lankton's model of patients' acceptance of e-health solutions [11]. The model was applied to measure behavioral intention to use (BI), intrinsic motivation (IM), perceived ease-of-use (PEOU), and perceived usefulness (PU) of the system.

BI denotes the intention to use and rely on the decision support system, IM denotes the readiness to use the decision support system without any compensation, PEOU denotes the

extent to which the generated reports are clear and understood by the patients, and PU refers to the degree to which the user believes that the use of the system would enhance their experience with laboratory tests.

We have applied the method proposed by Davis et al. in the revision of Wilson and Lankton [11] to measure BI, PEOU, and PU. IM was measured using the method provided by Davis et al. [12].

BI measure contained of 2 items whereas IM, PEOU, and PU contained 3 items each. We used Russian translations of the measurement tools made by the research team. To rate each item a Likert scale from 1 (not at all) to 7 (very much) was applied [13].

Recruitment

Participants were recruited in Saint-Petersburg, Russia. The inclusion criteria were: having experience with the decision support system with at least 5 reports on the test results.

Demographic details of the patients are presented in the table 1. IT literacy of the patients was evaluated based on how often they use PC or Smartphones. The IT literacy was graded from beginners – users that have started using PC within 6 months; intermediate – users of PC at least twice a week; and advanced – for those using PC on the daily basis. In total we have recruited 60 patients. Every of them has participated in the study.

Table 1 – Demographic details of the patients

Gender	Average age	Education					IT habits		
		Age > 60	Higher	Secondary	Below secondary	Beginners	Intermediate	Advanced	
28 Males	41.3	8	9	14	5	3	17	8	
32 Females	42.3	6	11	18	3	8	4	0	
Total 60	41.9	14	20	32	8	11	21	8	

Data collection and analysis

The participants were asked to complete an online questionnaire. Each participant got an individual link and could work only with one questionnaire. A detailed instruction on how to work with a questionnaire and the meaning of the scale was provided to each participant.

GNU Octave was used to calculate statistics of the participants' general characteristics and user acceptance measurements.

The study got the approval by the regional ethics committee. All the participants invited to the study were notified of the objectives of the study and of the purpose of the questionnaires. Each participant was assured, in writing, of their entitlement to anonymity and confidentiality. Written consent was acquired from each participant and each participant was reminded of their entitlement to withdraw data from the study database for up to three months after their approval.

Results

Correctness

As described in the methods section a sample of 256 reports was independently reviewed by two laboratory physicians of the laboratory service. The results of evaluation (e.g. precision, recall and F-measure) for each are presented in table 2. Cohen's kappa was calculated to check the inter-rater agreement between the two laboratory physicians. The physicians showed disagreement in the case of 2 reports.

Table 2 – Reports' quality evaluation

Mistakes	Precision	Recall	F-measure	Cohen's kappa
7	0.99	0.99	0.99	0.99

A 95%-confidence interval of 7 mistakes in 256 samples was calculated. The values is 0.8% - 5.1%.

Acceptance

Results of acceptance survey are presented in the Table 3.

Table 3 – User acceptance measures

Criterion, Item	Mean value	Standard deviation	Reference range
Behavioral intention to use	5.9	0.42	4.51-7.24
I intend to use the reports	5.7	0.43	4.46-6.91
I feel like I will use it in the future	6.1	0.41	4.78-6.57
Intrinsic motivation	6.2	0.48	4.98-7.41
I find the system to be useful for me	6.2	0.48	4.98-7.41
The system helps me to make more informed decisions	5.9	0.47	4.52-7.17
The system is reliable and I trust it	6.4	0.48	5.28-7.81
Perceived ease of use	5.7	0.50	4.32-7.02
The reports are clear and understandable	6.3	0.52	5.21-6.79
It is easy to access the reports	5.7	0.48	4.28-7.04
I like that I can keep all my reports in the electronic format	5.4	0.49	4.78-6.78
Perceived usefulness	5.9	0.46	4.48-7.15
Using the system enhances the effectiveness of managing my health conditions	5.7	0.48	4.28-7.04
It explains me what my health status is	6.1	0.44	4.54-6.93
I can provide all the information about my test results to any doctor I visit	5.9	0.46	4.52-7.19

The mean values for BI, IM, PEOU, and PU (5.9, 6.2, 5.7, and 5.9 respectively) showed the high acceptance of the decision support system and the reports, which it generates.

Discussion

The evaluation showed that the correctness of the generated reports is high. Seven mistakes out of 256 analyzed reports were caused by a human factor. The value of 5.1% of the 95% confidence interval for the 256 items sample demonstrates that even a small amount of mistakes lead to the problem of correct reports generation. To solve the problem of correct rule definition we have identified mistakes in the rules, that were defined by the experts. Out of 151 rules defined by the experts only 2 resulted in the faulty reports. All of them were

caused by typos that could be easily avoided. This led to a change of rules' definition procedure, where we apply 4 eyes principle now. This means that each rule shall be reviewed and approved by a second expert.

The user acceptance of the system was evaluated after two months of operation. Acceptance scores were relatively high. Unfortunately, we could not compare them to the similar studies, as we did not find a patient oriented decision support system, for which a user acceptance was evaluated.

Current research is focused on the extension of the knowledge representation language by adding an ability to work with fuzzy sets [7]. This will provide experts with flexibility in definition of knowledge and rules. We also are studying the possibility to validate the reports that are produced by DoctorEase to enable the system acquiring knowledge based on its experience using case based reasoning approach.

For the patients we are developing a mobile application to make the work with the system easier.

Conclusion

The paper presents an evaluation of a decision support system for the patients. The results show that the system is demanded and trusted by the patients. And this approach is a of a high demand especially in the situation when a patient can not refer to a doctor to get a professional interpretation of the results.

Acknowledgements

The research was supported by the Grant of a Russian President #AAAA-A16-116120810057-8.

References

- [1] L.J. Johnson, Tests, missed appointments require you to follow up, *Med Econ* **88** (2011), 114-115.
- [2] I.E. Smith and G. Schiavon, Follow-up tests to detect recurrent disease: patient's reassurance or medical need?, *Breast* **22 Suppl 2** (2013), S156-160.
- [3] R. Haeckel, W. Wosniok, and F. Arzideh, Proposed classification of various limit values (guide values) used in assisting the interpretation of quantitative laboratory test results, *Clin Chem Lab Med* **47** (2009), 494-497.
- [4] L. Ahmadian, M. van Engen-Verheul, F. Bakhshi-Raiez, N. Peek, R. Cornet, and N.F. de Keizer, The role of standardized data and terminological systems in computerized clinical decision support systems: literature review and survey, *Int J Med Inform* **80** (2011), 81-93.
- [5] G. Kopanitsa, Standard based multiclient medical data visualization, *Stud Health Technol Inform* **180** (2012), 199-203.
- [6] G. Kopanitsa, Evaluation study for a multi-user oriented medical data visualization method, *Stud Health Technol Inform* **200** (2014), 158-160.
- [7] G. Kopanitsa, Z. Tsvetkova, and H. Veseli, Analysis of metrics for the usability evaluation of EHR management systems, *Stud Health Technol Inform* **180** (2012), 358-362.
- [8] G. Kopanitsa, Z. Tsvetkova, and H. Veseli, Analysis of metrics for the usability evaluation of electronic health record systems, *Stud Health Technol Inform* **174** (2012), 129-133.
- [9] T. Kawada, Sample size in receiver-operating characteristic (ROC) curve analysis, *Circ J* **76** (2012), 768; author reply 769.
- [10] K.J. Berry, J.E. Johnston, and P.W. Mielke, Jr., Weighted kappa for multiple raters, *Percept Mot Skills* **107** (2008), 837-848.
- [11] E.V. Wilson and N.K. Lankton, Modeling patients' acceptance of provider-delivered e-health, *J Am Med Inform Assoc* **11** (2004), 241-248.
- [12] S.A. Snipes, F.A. Montiel-Ishino, J.M. Smyth, D.J. Murphy, P.Y. Miranda, and L.A. Davis, User Perceptions of inverted exclamation markProtejasel!: An Intervention Designed to Increase Protective Equipment Use Among Mexican Immigrant and Mexican American Farmworkers, *JMIR Mhealth Uhealth* **4** (2016), e28.

- [13] B.L. Drinkwater, A comparison of the direction-of-perception technique with the Likert method in the measurement of attitudes, *J Soc Psychol* **67** (1965), 189-196.

Address for correspondence

Georgy Kopanitsa. Institute Cybernetic Center, Tomsk Polytechnic University. 634050 Lenina 30, Tomsk, Russia.
Email: georgy.kopanitsa@gmail.com

Effective Usability Engineering in Healthcare: A Vision of Usable and Safer Healthcare IT

Andre Kushniruk^{a,b}, Yalini Senathirajah^c, Elizabeth Borycki^a

^aSchool of Health Information Science, University of Victoria, Canada,

^bDepartment of Development and Planning, Aalborg University, Aalborg, Denmark

^cState University of New York, Brooklyn, USA

Abstract

Persistent problems with healthcare IT that is unusable and unsafe have been reported worldwide. In this paper we present our vision for deploying usability engineering in healthcare in a more substantive way in order to improve the current situation. The argument will be made that stronger and more substantial efforts need to be made to bring multiple usability engineering methods to bear on points in both system design and deployment (and not just as a one-time effort restricted to software product development). In addition, improved processes for ensuring the usability of commercial vendor-based systems being implemented in healthcare organizations need to be addressed. A discussion will also be provided on challenges and barriers that will need to be overcome to ensure that the healthcare IT that is released is both usable and safe.

Keywords:

User-Computer Interface; Patient Safety; Health Information Systems

Introduction

Usability of healthcare IT has continued to be a worldwide issue, with continued reports of systems that are unusable, negatively affect healthcare workflow, and might even introduce a new class of error – technology-induced error [1-3]. There are a wide range of problems being reported that are related to human-computer interaction, including poor usability; the inability to customize systems to local needs, terminologies and workflows; and problems integrating information from multiple systems [4-7]. In response, there has been an attempt to create more effective processes for developing healthcare IT systems, including the introduction of certification processes in an attempt to encourage and mandate user-centered design and more usable systems.

In the United States, the Office of the National Coordinator (ONC) has developed guidelines for certifying the usability of vendor-based healthcare IT, including electronic health records. These have included vendors showing evidence of having applied user-centered design processes and usability testing in the design process [8]. In Europe, similar processes are also beginning to be mandated with CE marking [9]. However, Ratwani and colleagues have shown that the application of user-centered design processes by electronic health record (EHR) vendors is quite variable, even when adhering to the new regulations [10]. Results of their work also indicated that 63% of 41 vendors studied used fewer than 15 participants in usability testing and only 9% used at least 15 participants with clinical backgrounds. Currently, ONC-authorized certification organizations certify EHR products, with the vendor being required to provide a written statement

about the process they used, along with results of usability tests. However, Ratwani and colleagues also found that there is a lack of adherence to certification requirements and standards (even among EHR products that were certified as meeting these requirements) [10]. Despite the importance of such certification as a first step, the current approaches are also limited in that they focus only on the system design/development process, and they do not extend to the implementation processes involved in system deployment of vendor systems in real world settings (such as hospitals and hospital systems). In such settings, customizing and modifying systems such as EHRs approaches the complexity of software product development. However, similar regulation covering practices for applying usability testing in the implementation of commercial vendor-based systems in real hospitals and clinical settings has been absent.

In this vision paper we argue for the need to distinguish between usability engineering needs in the: (a) one-time design process of systems (where certification efforts have been aimed so far), and (b) continual and long-term need for usability engineering in the implementation, customization and re-implementation of healthcare IT products. In addition, the argument will be made that a multi-level approach, borrowed from work in human-computer interaction in healthcare, is needed for ensuring healthcare IT usability and safety.

Need for Improved Usability Engineering in Healthcare IT Product Development

A variety of approaches have been developed and used for testing the usability of systems and assessing their impact on patient safety. However, as noted above, currently there are continued reports of a lack of system usability [5-7]. One of the issues is that the current certification requirements are limited and do not take into account the need for a variety of different approaches to applying usability engineering methods in the context of the multiple levels of complexity in healthcare. Along these lines, the context of use is critical to consider when conducting usability analyses. The testing of systems outside of the context of their application can lead to issues when the systems are actually used, and so an argument has been made for improved usability engineering processes. However, with a centralized product certification approach system usability and safety is assessed outside of the context of the healthcare organization where it will ultimately be used. This is problematic when systems are later released into varied contexts within a region, across a country, and particularly when systems developed and certified in one country are imported into another country, where the healthcare system and workflow may be very different from the country in which the system was developed. Greater emphasis on testing systems in a variety of local contexts is also needed before releasing new healthcare IT products.

Need for Usability Engineering in Healthcare IT System Implementation and Customization

As noted above, usability certification processes are beginning to address the issue of a lack of usability of IT products. However, it should be noted that the centralized certification of vendor-based systems, such as EHRs, by regulatory bodies does not guarantee that such systems (post development) will either be usable or free from technology-induced error when later implemented in hospitals and healthcare settings. Along these lines, in this paper we also argue for the development of best practices for usability processes not only in system design and development, but also in the implementation of vendor-based EHR products in healthcare settings. This will be needed in order to ensure the entire process of implementation includes appropriate consideration for usability and safety. In our vision, which is based on advances in human-computer interaction, this would include a multi-level approach to testing and customizing commercial systems being implemented in healthcare settings.

Towards a Layered Approach to Usability Engineering Throughout the SDLC

Work in usability engineering in healthcare has shown that the application of methods emerging from the field of human-computer interaction need to be considered at multiple levels. Given the complexity of healthcare, it is not enough to test systems for human-computer interaction in isolation of their real use in complex and dynamic settings. Typical usability testing (such as it is specified currently by certification bodies) consists of observing and recording a small number of end users (e.g. physicians or nurses) as they interact with the system to carry out representative tasks (e.g. entering medication orders into a computer system). This level (which we refer to as Level 1) is useful to help identify and screen off surface level usability problems (such as labelling problems, navigational issues, font size issues, etc.). However, as shown in Figure 1, in order to ensure that healthcare IT will work properly in the context of carrying out work activities involving complex cognitive and social processes, Level 1 testing is not sufficient, and additional layers of testing will be required. (See Figure 1 for the proposed levels and their associated usability engineering methods listed on the right hand side of the figure.) As an example, Level 1 usability testing might involve observing physicians interacting with an EHR in isolation to carry out tasks and might determine that specific surface level aspects of the user interface need to be improved (such as making alerts and reminders more prominent on the computer screen). Although this might satisfy current certification requirements during the product's development, prior to releasing the system in hospitals it may also be important to know if the system integrates into daily work practices. For example, in the case of an EHR it might be important to know whether or not the system can be easily used

during actual clinical interactions or if it interferes with clinical reasoning. In order to more fully test systems, a second level of testing is then needed: clinical simulations, as illustrated as Level 2 in Figure 1. (The up arrow in the figure shows the typical recommended progression of testing from Level 1 up to Levels 2 and 3.)

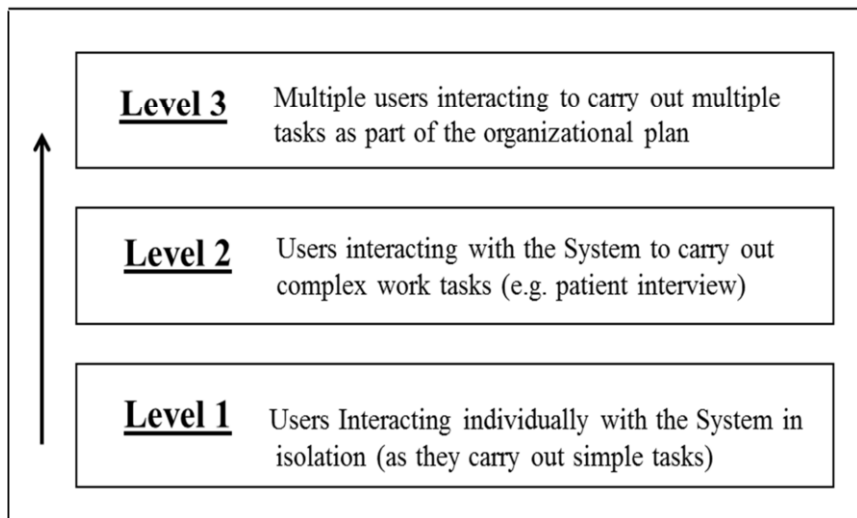
Clinical simulations extend usability testing by examining systems under real or realistic conditions, settings and contexts of use. For example, in order to assess if an EHR system works as expected when a physician user interacts with it during a patient interview, a clinical simulation can be created whereby the user's (i.e. physician's) interactions with the system are recorded while they interact with either a real or standardized patient (i.e. someone playing the role of the patient). Such testing, although essential to ensure a system works in realistic contexts of use, is beyond the scope of current regulations or certifications. However, a variety of published studies have reported on how this type of clinical simulation can be set up with relatively low cost and high impact for improving the design, as well as customization, of EHR systems [19].

Although clinical simulations are useful in assessing the potential impact of systems on workflow and more complex cognitive activities of health professionals, they can never fully predict how a system will work under real conditions. Along these lines, additional testing in restricted live or near-live contexts of use is recommended (see Level 3 in Figure 1). Studies where this was done have indicated it was worth the effort in terms of reducing usability problems and avoiding errors [11].

What Will be Required for this Improved Application of Usability Engineering in Healthcare?

Currently, a variety of usability methods and approaches exist that have been applied in projects and published in the health informatics literature. For example, work has been conducted in carrying out a wide range of usability tests, clinical simulations, and combinations of these approaches in order to improve healthcare IT usability and safety [19]. This has included work in: (a) improvement of software, (b) improvement of the process of development, (c) customization of vendor products and improvement in user training, (d) new approaches to software testing, and (e) selection and procurement of safe healthcare IT [12-20]. Along these lines, Marcilly and colleagues have argued for "evidence-based usability" in health IT, as the scientific body of reported methods that have proven effective has grown considerably [21]. Some of the results and methods that have emerged from this work have begun to be incorporated into regulatory processes in a number of countries [22]; however, progress in doing this has not kept up with the increasing reports of issues related to usability and safety of healthcare IT.

Levels of System Testing



Testing Methods

Clinical Simulations & Naturalistic Study (detects organizational impact)

Clinical Simulation (detects impact on workflow and cognition)

Usability Testing (detects surface usability problems)

Figure 1 - A Multi-layered Approach to Usability Engineering in Healthcare

Despite the research in this area, healthcare is still plagued by unusable and potentially unsafe systems. So the question remains as to what is needed for a vision of more usable and safer systems, and how can that vision be attained?

Recommendations and Potential Directions

In order to achieve usable and safe healthcare IT, a number of recommendations and potential directions for further work and effort have emerged from the earlier discussion in this paper and are presented below.

Firstly, regarding more thorough usability evaluations, it is argued that all 3 levels in Figure 1 need to be considered when (a) developing health information technology or systems and (b) when customizing and deploying complex systems (such as EHRs) in real hospital and healthcare environments. It has been found that problems detected at Level 1 do not encompass the full range of usability issues. This will require going beyond current usability certifications and regulations to include multi-level usability evaluation. It will also require going beyond considering usability in the realm of system design and, as importantly, as part of organizational system deployment and implementation strategies. It will be essential to take into consideration the impact of local context when implementing healthcare IT.

Secondly, this vision will require certification and regulation with “teeth” that will involve more rigorous testing and lead to improved support and enforcement of basic usability principles. Along these lines, improved reporting mechanisms for the reporting of known usability and safety issues across the healthcare industry is needed globally (particularly as some systems are beginning to have a global market) [23-24]. Stronger regulatory measures will also be needed, specifying more stringent application of usability engineering approaches before certifying products.

Thirdly, a critical area that needs to be considered in making the vision of usable and safe systems a reality is the need for improved education about not only the importance of usability engineering and user-centered design, but also education about the full range of methods themselves that are included under these headings. This includes disseminating knowledge of alternative and emerging approaches at the levels of software designers, managers, clinical users and healthcare decision makers.

Fourthly, a critical issue that to some extent may prevent use of methods described in this vision paper is that of perceived cost and the amount of time it would take to use and apply methods known to lead to improved system usability and safety. Along these lines, there is growing evidence from studies on the cost-effectiveness of applying more thorough and advanced usability engineering processes [25]. These findings need to be more widely disseminated to not only healthcare IT professionals, but also to government and hospital decision makers, particularly when it comes to system procurement. Some progress has been made in the recent application of usability engineering leading to improved healthcare IT procurement by requiring test installs of systems and having them undergo rigorous usability evaluation before selecting software products for implementation [26].

Conclusion

A vision of healthcare IT can be conceived of where systems are seen as being highly usable; flexible within complex work activities; error-reducing, and sensitive to the contexts of different users, uses and locations. Unfortunately, reports continue to appear that seem to indicate that this ideal is not being met in many healthcare institutions [27]. To make the vision a reality will require considerably more effort along several fronts, including better dissemination of methods and

approaches that are known to work to improve healthcare IT usability, along with better sharing of this information and knowledge through improved practices and regulation. Furthermore, it should be the responsibility of healthcare organizations that buy and deploy healthcare IT to demand products that are designed to be more usable and that their IT staff be familiar with methods in usability engineering. From the vendor side, it is hoped that this could be achieved prior to extensive regulatory requirements, and improved usability and safety would become features of commercial products that provide market advantage (by distinguishing themselves in terms of improved user interaction and user experience). In summary, it has become clear that the usability of healthcare IT has become a critical issue that must be given greater consideration at multiple levels, from healthcare IT professionals to the governmental level. In this vision paper we have discussed approaches and methods we feel will be important in improving the current situation regarding usability and safety of healthcare IT.

Acknowledgements

This work is supported by the Agency for Healthcare Research and Quality (AHRQ) grant R01HS023708.

References

- [1] E.M. Borycki and E. Keay, Methods to assess the safety of health information systems, *Health Q* **13** (2010), 47-52.
- [2] A.W. Kushniruk, C. Nohr, S. Jensen, and E.M. Borycki, From Usability Testing to Clinical Simulations: Bringing Context into the Design and Evaluation of Usable and Safe Health Information Technologies. Contribution of the IMIA Human Factors Engineering for Healthcare Informatics Working Group, *Yearb Med Inform*, **8** (2012), 78-85.
- [3] A.W. Kushniruk, E.M. Borycki, S. Kuwata, and J. Kannry, Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments. *Stud Health Technol Inform* **169** (2010), 915-919.
- [4] B. Middleton, M. Bloomrosen, M.A. Dente, B. Hashmat, R. Koppel, J.M. Overhage, T. H. Payne, S.T. Rosenbloom, C. Weaver, and J. Zhang, Enhancing patient safety and quality of care by improving the usability of electronic health record systems, *J Am Med Inform Assoc* **20**(e1) (2013), e2-8.
- [5] A.L. Kellermann, and S.S. Jones, What it will take to achieve the as-yet-unfulfilled promises of health information technology, *Health Affairs* **32**(1) (2013), 63-68.
- [6] E. Jamoom, P. Beatty, A. Bercovitz, D. Woodwell, K. Palso, and E. Rechtsteiner, Physician adoption of electronic health record systems: United States, 2011. NCHS data brief, no. 98. Hyattsville, MD. National Center for Health Statistics, 2012.
- [7] L. Riskin, R. Koppel, and D. Riskin, Re-examining health IT policy: what will it take to derive value from our investment? *J Am Med Inform Assoc* **22**(2) (2015), 459-464.
- [8] ONC Meaningful Use and Usability Testing, <http://www.healthcareusability.com/article/onc-meaningful-use-and-usability-testing>, accessed September 20, 2016
- [9] Medical Device Directive MDD93/42/EEC, http://www.mdss.com/download/download_station.htm, accessed December 21, 201
- [10] R.M. Ratwani, N.C. Benda, A.Kettinger, and R. Fairbanks, Electronic health record vendor adherence to usability certification requirements and testing standards, *J Med Inform Assoc* **314**(10) (2016), 1070-1071.
- [11] A.C. Li, J.L. Kannry, A. Kushniruk, D. Chrimes, T.G. McGinn, D. Edo-nyabo, and D.M. Mann, Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support, *Int J Med Inform* **81**(11) (2012), 761-72.
- [12] M.C. Beuscart-Zéphir, S. Pelayo, F. Anceaux, J.J. Meaux, M. Degroisse, and P. Degoulet, Impact of CPOE on doctor-nurse cooperation for the medication ordering and administration process, *Int J Med Inform* **74**(7) (2005), 629-41.
- [13] M. C. Beuscart-Zéphir, P. Elkin, S. Pelayo, and R. Beuscart, The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges, *Yearb Med Inform* (2006), 109-27.
- [14] J. Zhang, T.R. Johnson, V. L. Patel, D.L. Paige, and T. Kubose, Using usability heuristics to evaluate patient safety of medical devices, *J Biomed Inform* **36**(1) (2003), 23-30.
- [15] D.R. Kaufman, V.L. Patel, C. Hilliman, P.C. Morin, J. Pevzner, R.S. Weinstock, R. Goland, S. Shea, and J. Starren, Usability in the real world: assessing medical information technologies in patients' homes, *J Biomed Inform* **36**(1) (2003), 45-60.
- [16] A.W. Kushniruk, and E.M. Borycki, Low-cost rapid usability engineering: designing and customizing usable healthcare information systems, *Health Q* **9**(4) (2006), 98-100.
- [17] M.W. Jaspers, T. Steen, C. van Den Bos, and M. Geenen, The think aloud method: a guide to user interface design, *Int J Med Inform* **73**(11) (2004), 781-95.
- [18] A.W. Kushniruk, E.M. Borycki, S. Kuwata, and J. Kannry, Predicting changes in workflow resulting from healthcare information systems, *Health Q* (2006), Oct 15;9(Sp).
- [19] E. Borycki, A. Kushniruk, C. Nohr, H. Takeda, S. Kuwata, C. Carvalho, M. Bainbridge, and J. Kannry, Usability methods for ensuring health information technology safety, *Yearb Med Inform* **8**(1) (2013), 20-27.
- [20] E. M. Borycki, and A.W. Kushniruk, Towards an integrative cognitive-socio-technical approach in health informatics: analyzing technology-induced error involving health information systems to improve patient safety, *The Open Medical Informatics Journal* **4**(1) (2010).
- [21] R. Marcilly, L. Peute, and M.C. Beuscart-Zéphir, From Usability Engineering to Evidence-based Usability in Health IT, In Evidence-Based Health Informatics: Promoting Safety and Efficiency Through Scientific Methods and Ethical Policy, 2016 May 20;222:126.
- [22] A.W. Kushniruk, D.W. Bates, M. Bainbridge, M.S. Househ, and E.M. Borycki, National efforts to improve health information system safety in Canada, the United States of America and England, *Int J Med Inform* **82**(5) (2013), 149-160.
- [23] F. Magrabi, M.S. Ong, W. Runciman, and E. Coiera, Using FDA reports to inform a classification for health information technology safety problems, *J Am Med Inform Assoc* **19**(1) (2012), 45-53.
- [24] S. Palojoki, T. Pajunen, K. Saranto, and L. Lehtonen, Electronic health record-related safety concerns: A cross-sectional survey of electronic health record users, *JMIR Med Inform* **4**(2) (2016), e13.
- [25] T.B. Baylis, A.W. Kushniruk, and E.M. Borycki, Low-cost rapid usability testing for health information systems: is it worth the effort? *Stud Health Technol Inform* **180** (2011) 363-367.
- [26] S. Jensen, S.L. Rasmussen, and K.M. Lyng, Evaluation of a clinical simulation-based assessment method for EHR-platforms, *Stud Health Technol Inform* **205** (2013) 925-929.
- [27] S. Palojoki, M. Mäkelä, L. Lehtonen, and K. Saranto, An analysis of electronic health record-related patient safety incidents. *Health Informatics Journal* (2016) Mar 7;1460458216631072.

Address for correspondence

Corresponding author: Dr. Andre Kushniruk, Professor and Director, School of Health Information Science, University of Victoria, Victoria, Canada, email: andrek@uvic.ca

Automated Classification of Multi-Labeled Patient Safety Reports: A Shift from Quantity to Quality Measure

Chen Liang ^a, Yang Gong ^b

^a Louisiana Tech University, Ruston, Louisiana, USA

^b School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas, USA

Abstract

Over the past two decades, there have seen an ever-increasing amount of patient safety reports yet the capacity of extracting useful information from the reports remains limited. Classification of patient safety reports is the first step of performing a downstream analysis. In practice, the manual review processes for classification are labor-intensive. Studies have shown that the reports are often mislabeled or unclassifiable based on the pre-defined categories, which presents a notable data quality problem. In this study, we investigated the multi-labeled nature of patient safety reports. We argue that understanding multi-labeled nature of reports is a key to disclose the complex relations between many components during the courses and development of medical errors. Accordingly, we developed automated multi-label text classifiers to process patient safety reports. The experiments demonstrated feasibility and efficiency of a combination of multi-label algorithms in the benchmark comparison. Grounded on our experiments and results, we provided suggestions on how to implement automated classification of patient safety reports in the clinical settings.

Keywords:

Patient Safety, Machine Learning

Introduction

As emphasized in the Institute of Medicine's report 'To Err Is Human', error reporting and analysis are fundamental to patient safety [1]. Classification of patient safety reports is recognized as an initial step before any analysis and interventions can be applied [2]. However, the task of classification presents two urgent problems. Firstly, classified patient safety reports are of low quality due to sizable mislabeled reports and the reports under miscellaneous categories [3]. This problem may be caused by the lack of a deep contextual understanding of the reports. Secondly, conventional report classification is labor-intensive and time-consuming. Presently, the classification task is largely completed by manual efforts. The rapid increase in the volume of the reports and research demands calls for an efficient and reliable solution for the classification task.

Patient safety reports are multi-labeled in nature. In many scenarios, an entity can be categorized with a single label. This is known as single-label classification. When an entity is associated with multiple labels, it is known as multi-label classification. Patient safety reports can be categorized with a number of pre-defined labels that fall into various categories such as incident types, type of harms, and contributing factors. The multi-labeled reports carry invaluable information for downstream analysis. Studies that evaluate patient safety reports

tend to survey types and frequency of errors but draw less attention in the co-occurrence and relation of labels. These pieces of information may provide critical insights of reasoning about the root causes. In a study, researchers discovered that 42% of the reports were associated with more than two contributing factors. The increasing number of contributing factors per report may indicate the variation, increased complexity or severity of harm [4]. The multi-labeled patient safety reports can draw reviewers' special attention and thus can be better understood if they are well organized by categories.

Automated classification can be an alternative to the situation yet it presents interesting challenges when it is adapted to multi-label tasks. Automated classification is the task of using computers to learn associations between examples and labels. In-patient safety reports, an example can be an individual report in a corpus of reports. The classification is supervised because a classifier is built and trained by a set of prior labeled reports. Once the classifier is well trained, it can be used to predict candidate labels for unlabeled reports. As opposed to single-label classification, each report in multi-label classification may be associated with one or more labels. The multi-label classification is intuitive in human cognitive processes but creates extra complications in the computational process. Firstly, the training of a classifier is affected by many factors including the co-occurrence frequency of labels, hierarchical label relations, etc. [5; 6]. Multi-label algorithms must take these pieces of information into account but can easily lose feasibility or scalability by introducing sharply increased computational complexity [7]. Secondly, label imbalance is nearly inevitable in a multi-labeled corpus. As such, it negatively influences both classification performance and the selection of evaluation metrics. In specific, the prediction power of minority labels will decline since reports that are associated with these labels are less weighted during the training phase. Evaluation metrics such as exact match may not be as sufficient as it is in single-label tasks since a report may be predicted partially correct.

In general, multi-label algorithms can be categorized into two approaches: problem transformation, and algorithm adaptation. Problem transformation methods transform the multi-label problems to a number of single-label problems where a single-label problem can be solved by a range of single-label algorithms. A list of well-documented problem transformation algorithms includes binary relevance (BR), pairwise classification (PW), label combination (LC), and ranking and threshold (RT) [6]. Recently, a pruned sets method (PS) that is adapted from BR is reported to overcome many drawbacks of BR such as computational complexity [8]. A classifier chain method (CC) is also reported to improve LC for its less consideration of label relations, and sizable computational com-

In this paper, we adapted multi-label classification methods to categorize multi-labeled patient safety reports. The problem transformation algorithms were employed as the method to transform multi-label tasks to single-label tasks where binary classification algorithms serve as base classifiers. We present benchmark comparison between different combinations of algorithms. We argue that patient safety reports are multi-labeled and can be categorized by automated classification method with advanced efficiency.

Patient safety reports contain crucial information to analyze incidents and to reduce recurrence. Reflected on a variety of factors mentioned in a report, an incident is multi-faceted. During the course of health care, a patient may have as many as hundreds of direct and indirect encounters with many persons, clinical procedures, and medical devices involved in hospital visits. Adverse outcomes may only occur in few encounters. But errors and contributing factors are likely to tie with many other encounters; especially those seemingly have no direct correlations to the patient harm. When an incident involves more than one error, the scenario is even more complicated. This situation is exactly the same as it is described in the Swiss cheese model of health care systems, which implies that medical errors are caused by simultaneous failures of multiple components in the system [10]. Fortunately, many of these factors can be extracted from the reports with either a clear mention or an indirect clue. Categorizing reports in a correct way has become a key to understanding both the root

When it comes to massive reports, exploring multiple factors will provide more insights into the system's failure. Assuming that each factor represents a label, there are one or more labels associated with each report. The frequency of each label across the corpus can provide a general view of the distribution of labels. Another view of label distribution is the co-occurrence probability, which measures the probability of co-occurred labels. The co-occurrence probability is a piece of critical information in disclosing label relations. From the clinical perspective, if there is a high co-occurrence of a medication ordering system, a documentation error, and various drug reactions, for example, it is likely to have a malfunction or reoccurred user error relating to the medication ordering system. In a certain sense, therefore, label co-occurrence partially implies the potential factors that correlate with a given event.

Figure 1 – Network visualization of co-occurrence probability of labels in 2,919 patient safety reports.

Automated Classification

In this section, we report our work to adapt multi-label text classification in the task of categorizing patient safety reports. The experiments were designed to evaluate the feasibility and efficiency of automated methods in multi-labeled reports.

Dataset

We used a corpus consisting of 2,919 de-identified patient safety reports from a university healthcare system. The original corpus contains 54 labels with a label cardinality of 2.89. The label cardinality is defined as

$$LCard(C) = \frac{\sum_{i=1}^N |y_i|}{N} \quad (1)$$

N denotes the number of reports in the corpus C , y is the number of labels associated to with individual report, and L denotes the total number of labels. To avoid extreme imbalance of labels, we removed minority labels where each has less than 50 reports, resulting in 28 labels with a label cardinality of 2.58. See Table 1.

Table 1 – A demonstration of 28 labels in a hierarchy.

Top Level Category	Label	Frequency
Incident Type	Behavior	170
Incident Type	Clinical Administration	253
Incident Type	Device	317
Incident Type	(6 more)	-
Error Type	Adverse Drug Reaction	258
Error Type	Failure/Malfuction	105
Error Type	Fall at Bed	110
Error Type	(11 more)	-
Harm	Injury	123
Harm	Suffering	233
Contributing Factor	Behavioral Factor	169
Contributing Factor	Communication Factor	256
Contributing Factor	Performance Factor	1513

Procedure

We used problem transformation methods to solve the multi-label challenge. The problem transformation methods require a problem transformation algorithm in conjunction with a single-label algorithm that serves to build base classifiers. To perform the benchmark comparison, we chose a number of well-documented problem transformation algorithms considering both feasibility and computational complexity including BR, LC, RT, CC, and PS. Note that PW is not selected due to its significant computational complexity. For the parameter selection for PS, we used the optimized parameters ($n = 0$; $p = \{1, 3\}$) given that our corpus has a label cardinality of 2.58 and a number of 28 labels [8]. We also chose a number of single-label classifiers that represent a range of well-developed algorithms consisted of Naïve Bayes [11], Support Vector Machine (SVM) [12], k-Nearest Neighbor (kNN) [13], Decision Tree [14], and Decision Rule [15]. See Table 2.

Since these binary classifiers are not originally designed to process text data, we prepared our corpus as follows. (1) Snowball stemmer was used to reduce inflected terms to their root form [16]. (2) Rainbow list was used to remove stop words [17]. (3) Alphabetic tokenizer was used to break a string of text into terms. (4) Lower case token was applied to all the terms. (5) TF-IDF (term frequency-inverse document frequency) was used in the transformation of documents into a bag-of-words (BOW) matrix keeping 1000 unique terms [18].

Table 2 Single-label classification algorithms used in the experiments.

Algorithm	Implementation	Parameter
Naïve Bayes	NaiveBayes	
Support Vector Machine	LibSVM	Linear SVM
k-Nearest Neighbor	IBk	$k = 1$
Decision Rule	JRip	
Decision Tree	J48	

The experiments were performed on a 64-bit OS system with a processing power of 2.2 GHz with 4 Cores 8 Threads and a memory of 8 GB RAM. We used Python 3.0 to pre-process the corpus. The model training, evaluation, and statistics were performed on WEKA 3.6 [19] and MEKA 1.9.0 [20]. We employed 5×2 fold cross validation to randomize data and average results.

Evaluation Metrics

We consider both essential evaluation measures that are used in the single-label classification and the ones that are adapted for multi-label classification. The 0/1 Loss is a loss measure that assigns a '1' only if a label set is predicted exactly correct. The 0/1 Loss is defined as

$$0/1 \text{ Loss} = \frac{1}{N} \sum_{i=1}^N 1_{(\hat{Y}_i \neq Y_i)} \quad (2)$$

where \hat{Y} denotes the predicted set of labels; Y denotes the exact set of labels.

Hamming Loss is the measure of labels that are incorrectly predicted. Instead of penalizing the incorrect match between two sets of labels, the Hamming Loss measures only the symmetrical difference between individual labels. Therefore, it is more forgiving than the 0/1 Loss. Hamming Loss can be referred as a partial match metrics whereas 0/1 Loss measures exact match. In multi-label classification, the form of Hamming Loss is defined as

$$\text{Hamming Loss} = \frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^L 1_{(\hat{y}_{i,j} \neq y_{i,j})} \quad (3)$$

We also employed a multi-label accuracy measure that has been widely used in multi-label classification [21]. The form is defined as

$$\text{Multi-label Accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i \cap \hat{y}_i|}{|y_i \cup \hat{y}_i|} \quad (4)$$

The F measure is employed to provide a balanced metric between precision and recall. In multi-label classification, the generic F measure can take a number of forms depending on the different approaches of averaging metrics. In this study, we employed micro F measure. Mathematically, micro F measure favors labels with more documents, as it counts global metrics where labels with more documents have more weights. It is defined as

$$F_{\text{Micro}} = F \text{ Measure}(y_{i,j}, \hat{y}_{i,j}) \quad (5)$$

In addition, we also consider the build time of models as a measure of computational complexity.

Results

We present the results by each evaluation metric. BR ($M = 0.994$) and LC ($M = 0.994$) performed slightly better than other multi-label algorithms but did not lead a clear difference in terms of 0/1 Loss. Naïve Bayes ($M = 0.990$) shows a slightly higher rate compared to other binary algorithms.

LC ($M = 0.141$) and PS_(p=3, n=0) ($M = 0.140$) reveal an overall better performance in the measure of Hamming Loss. SVM ($M = 0.136$) shows the best performance among binary classifiers. The best combination appears to be the SVM in conjunction with RT (0.115).

BR ($M = 0.110$) is shown as the best problem transformation algorithm in the measure of multi-label accuracy. In the comparison of the base classifiers, Naïve Bayes ($M = 0.110$) outperformed the others. Naïve Bayes in conjunction with CC (0.128) appears to be the best combination.

BR ($M = 0.195$) reveals a higher F than other problem transformation algorithms. Naïve Bayes ($M = 0.190$) is the best base classifier overall. The best combination is Naïve Bayes in conjunction with CC (0.222). See Table 3 for details.

Table 3 – Micro F measure for different multi-label classifiers.

	NB	SVM	kNN	DR	DT	Rank
BR	0.212	0.184	0.152	0.219	0.206	1
LC	0.173	0.166	0.088	0.101	0.180	4
RT	0.185	0.118	0.122	0.190	0.192	3
CC	*0.222	0.197	0.092	0.209	0.165	2
PS (p=1, n=0)	0.173	0.165	0.087	0.104	0.163	5
PS (p=3, n=0)	0.173	0.163	0.084	0.111	0.157	6
Rank	1	3	5	4	2	

* best performance

PS_(p=3, n=0) ($M = 10.267$ second) is the most efficient problem transformation algorithm as it shows the shortest build time. In terms of base classifiers, Naïve Bayes, SVM, and kNN ($M = 0.228$ second) cost least build time. kNN in conjunction with PS_(p=1, n=0) (0.015 second) is the most efficient combination.

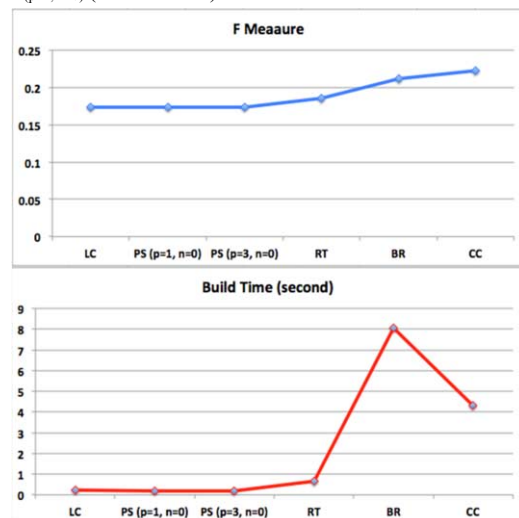


Figure 2 – F measure and build time for different problem transformation algorithms in conjunction with Naïve Bayesian.

We observed an interactive effect between F measure and build time over problem transformation algorithms. Figure 2 shows the comparison between different problem transformation algorithms when the base classifier is Naïve Bayes. CC appears to be a well-balanced algorithm considering its optimal performance and relatively short build time. Such an interactive effect did not show over binary algorithms.

Discussion

Multi-labeled Reports: from Quantity to Quality

This paper provided a novel perspective to investigate the nature of patient safety reports that is a report can be categorized by multiple labels. This argument corresponds to the Swiss cheese model that is frequently used as a metaphor of health care systems [10]. Errors are multi-faceted if they are viewed in a health care system where systems failures are preventable based on a number of relatively independent components. Such a component may refer to clinical administration, treatments, uses of medical equipment, etc., while each may produce unintended consequences. A systems failure occurs when by chance all the components produce errors. However, a systems view of these components can be obtained only if the reports are well categorized and, consequently, an aggregate analysis is made available.

The multi-labeled perspective of reports motivates a shift from quantity to quality measure. Patient safety studies have made remarkable progress, notably in constructing a nationwide reporting mechanism and ongoing focus to reducing harm by learning from lessons. These efforts have largely increased the number of reports, establishing a quantitative measure of patient safety such as the distribution of occurrences and frequency of errors. In fact, we are still far away from timely analysis and targeted quality improvement implied by the event reports. The vulnerability of health care system calls for special attention. We argue that one crucial gap is the limited understanding of reports, especially the intricate relations of the factors involved in a report. When the volume of reports increases, such pieces of relational information become more robust to indicate systems vulnerability. A prerequisite of performing quality measure as such is the capacity of extracting complex factors from massive reports. In this study, we suggested a multi-labeled approach.

Clinical Implementation

Our findings hold promise to improve the large-scale classification of patient safety reports. The experiments suggest feasibility and efficiency of using automated multi-label classification method to categorize patient safety reports. To balance the predictive power and efficiency, we found that CC in conjunction with Naïve Bayes is well performed. In addition, PS is also a promising method, as it largely reduces the model build time within a relatively small decline of predictive power.

More importantly, the multi-label classification is suited to the existing event reporting systems for improving the existing single-labeled classification and manual procedures. Clinicians who are responsible for case review and aggregate data analysis are expected to benefit from the automated classification. However, note that automated classification may not completely replace human effort for two reasons. Firstly, the automated classification results are not expected as good as manual results. The automated classification results should

serve as a reference source during the human review processes. Secondly, the automated classification has a limited capacity of predicting rare cases since it is generated based on the existing data. The rare cases, instead, may provide unique or crucial insights into medical errors.

With the scope of implementing automated classification in the reporting systems, we further suggest a uniformed classification scheme, which should provide a consistent and up-to-date classification hierarchy across health care providers. This is important to automated classification because the classification is supervised in a way relying on predefined labels to predict unlabeled reports. In the U.S., the Common Formats are widely used as a reporting guideline and classification schema in nationwide reporting. To extend the Common Formats' influence in guiding multi-label classification, additional work must be done to develop a classification hierarchy that supports automated classification directly, such as developing an ontological representation of patient safety reports [22].

Limitation and Future Direction

The label imbalance problem is a limitation, particularly in the context of multi-label classification. However, the label imbalance is inevitable in multi-label classification. The minority labels are even more common in medical corpus because medical entities (e.g., diseases, phenotypes, etc.) are not evenly distributed in a population. For example, minority labels can be 'death' or 'performance factor' in our corpus. To partially solve the problem, we removed extremely biased labels instead of creating synthesized reports. Consequently, it may lose some labels that are clinically important. Therefore, our approach still needs human guidance on categorizing minority labels at the current stage. In the future, we plan to enrich the corpus in both volume and sources.

We also note that the overall predictive power is comparatively small. Partially because we did not choose to perform feature selection and other manipulations that ought to boost the performance, as well as the fact that the 5×2 fold cross validation purports to find the most competitive classifiers under limited resources. The other interpretations may be the intricate semantic information and loss of information in the real-world patient safety reports. In the next step, we will investigate the effects of semantic information and domain knowledge in classification tasks.

Conclusion

The study demonstrated the effectiveness and efficiency of using automated multi-label classification on real-world patient safety reports. Our findings may improve (1) the process of understanding medical errors from an aggregate analysis, (2) clinical implementation of automated classification for large-scale patient safety reports.

Acknowledgements

The study is supported by a research grant (1R01HS022895) from the Agency of Healthcare Research and Quality, and the University of Texas System Grants Program (#156374).

References

- [1] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, *To Err Is Human: Building a Safer Health System*, National Academies Press, 2000.

- [2] L.L. Leape and S. Abookire, WHO draft guidelines for adverse event reporting and learning systems: from information to action, (2005).
- [3] Y. Gong, Data consistency in a voluntary medical incident reporting system, *J Med Syst* **35** (2011), 609-615.
- [4] P.J. Pronovost, D.A. Thompson, C.G. Holzmueller, L.H. Lubomski, T. Dorman, F. Dickman, M. Fahey, D.M. Steinwachs, L. Engineer, J.B. Sexton, and others, Toward learning from patient safety reporting systems, *Journal of critical care* **21** (2006), 305-315.
- [5] Y. Labrou and T. Finin, Yahoo! as an ontology: using Yahoo! categories to describe documents, in: *Proceedings of the eighth international conference on Information and knowledge management*, 1999, pp. 180-187.
- [6] G. Tsoumakas and I. Katakis, Multi-label classification: An overview, *Dept. of Informatics, Aristotle University of Thessaloniki, Greece* (2006).
- [7] J. Read, Scalable multi-label classification, (2010).
- [8] J. Read, B. Pfahringer, and G. Holmes, Multi-label classification using ensembles of pruned sets, in: *2008 Eighth IEEE International Conference on Data Mining*, 2008, pp. 995-1000.
- [9] J. Read, B. Pfahringer, G. Holmes, and E. Frank, Classifier chains for multi-label classification, *Machine learning* **85** (2011), 333-359.
- [10] J. Reason, Human error: models and management, *BMJ* **320** (2000), 768-770.
- [11] G.H. John and P. Langley, Estimating continuous distributions in Bayesian classifiers, in: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 338-345.
- [12] C.C. Chang and C.J. Lin, LIBSVM: a library for support vector machines, in: *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, 2011, p. 27.
- [13] J. Wang and J.D. Zucker, Solving multiple-instance problem: A lazy learning approach, in: *Proc. 17th International Conf. on Machine Learning*, Morgan Kaufmann, 2000.
- [14] J.R. Quinlan, Induction of decision trees, *Machine learning* **1** (1986), 81-106.
- [15] W.W. Cohen, Fast effective rule induction, in: *Proceedings of the twelfth international conference on machine learning*, 1995, pp. 115-123.
- [16] M.F. Porter, Snowball: A language for stemming algorithms, in, 2001.
- [17] A. McCallum, Rainbow project, 1998.
- [18] G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information processing & management* **24** (1988), 513-523.
- [19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, The WEKA data mining software: an update, *ACM SIGKDD explorations newsletter* **11** (2009), 10-18.
- [20] J. Read, P. Reutemann, B. Pfahringer, and G. Holmes, Meka: a multi-label/multi-target extension to weka, *Journal of Machine Learning Research* **17** (2016), 1-5.
- [21] S. Godbole and S. Sarawagi, Discriminative methods for multi-labeled classification, in: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2004, pp. 22-30.
- [22] C. Liang and Y. Gong, Knowledge Representation in Patient Safety Reporting: An Ontological Approach, *Journal of Data and Information Science* **1** (2016), 75-91.

Address for correspondence

Yang Gong, MD, PhD; Yang.Gong@uth.tmc.edu; 713 500 3547
7000 Fannin St. Suite 600, Houston, Texas, 77030, USA

Predicting Harm Scores from Patient Safety Event Reports

Chen Liang^a, Yang Gong^b

^a Louisiana Tech University, Ruston, Louisiana, USA

^b School of Biomedical Informatics, University of Texas Health Science Center, Houston, Texas, USA

Abstract

The identification of the severity of patient safety events promotes prioritized safety analysis and intervention. The Harm Scale developed by the Agency for Healthcare Research and Quality is widely used in the US hospitals. However, recent studies have indicated a moderate to poor inter-rater reliability of the Harm Scale across a number of US hospitals. Although the reasons are multi-folded, biased human judgments are recognized as a prominent factor. We proposed that key information to identify and refine the severity of harm is contained in the narrative data in patient safety reports. Using automated text classification to categorize harm scores is intended to provide reduced subjective judgments and much improved efficiency. We evaluated different types of classification algorithms using a corpus of patient safety reports from a US health care system. The results demonstrate the effectiveness and efficiency of the proposed methods. Accordingly, human biases on the application of harm scores are expected to be largely reduced. Our finding holds promise to serve as a semi-supervised tool during the process of manually reviewing and analyzing patient safety events.

Keywords:

Patient Safety, Patient Harm, Data Mining

Introduction

Harm Classification and Scales

Reducing patient harm is a top priority of US hospitals and health care organizations. During the past two decades, researchers have been focusing on compiling patient safety events and detecting errors through nationwide patient safety reporting [1,2]. Event reporting at all levels has shown remarkable advantages to gather concurrent and retrospective events, including patient harms, near misses, and unsafe conditions in a timely-manner. Most importantly, it enables a close analysis on aggregate data, which increases the chance of disclosing vulnerability of health care systems. To accommodate safety event reporting at federal level, the Patient Safety Organization (PSO) has employed standardized event reporting formats (a.k.a., the Common Formats) to collect and classify reports [3].

Severity of safety events is an influential factor that can be identified by using the Common Formats. This piece of information plays a crucial role in triggering intervention actions and prioritizing limited resources of root cause analysis. In the Common Formats, Harm Scale is used to describe the degree of harm by assigning each event a harm score. The latest version of Harm Scale (v1.2) released in 2012 consists of a 5-point scale of severity of harm and a 2-point scale of anticipated duration of the harm (see Figure 1). In the meantime, a number of health care

organizations have also participated in developing harm scales from different perspectives. The World Health Organization (WHO) developed a five-point harm scale, consisting of 'no harm to death', 'mild', 'moderate', 'severe', and 'death' [4]. This scale is centered on the patient harms arising from the provision of care. The National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) developed a scale that takes duration and permanency of harm into account. The Institute for Healthcare Improvement (IHI) also developed Global Trigger Tools to measure the severity of patient harm [5].

7. After any intervention to reduce harm, what was the degree of residual harm to the patient from the incident (and subsequent intervention)? CHECK FIRST APPLICABLE:

AHRQ Harm Scale

a. ☐ Death: Dead at time of assessment. → ANSWER QUESTION 9

b. ☐ Severe harm: Bodily or psychological injury (including pain or disfigurement) that interferes significantly with functional ability or quality of life.

c. ☐ Moderate harm: Bodily or psychological injury adversely affecting functional ability or quality of life, but not at the level of severe harm.

d. ☐ Mild harm: Bodily or psychological injury resulting in minimal symptoms or loss of function, or injury limited to additional treatment, monitoring, and/or increased length of stay.

e. ☐ No harm: Event reached patient, but no harm was evident. → ANSWER QUESTION 8

f. ☐ Unknown

Figure 1 – A Harm Scale Screenshot from the Common Formats, Patient Information Form (v1.2).

Harm Scale Reliability

One of the most significant challenges of using harm scale is reliability, which is the deviation between reporters' judgment about the type and severity of harm [6-8]. In practice, the deviation not only influences the classification of harm but also relates to the determination of intervention actions. For example, if an event is determined at a certain level that is preventable, it is likely that significant analysis and intervention efforts will be assigned. Otherwise, the complication of the care occurred in the event is likely to be labeled as a risk factor.

The reliability of harm scale, especially standard scales used at a national level, is not as high as expected in the practice. A recent survey study on the reliability of the Common Formats Harm Scale across different roles of clinicians and different settings yielded an overall moderate level of reliability [8]. The findings in this study show that some levels of harm are difficult to distinguish from neighbor levels. This problem is most significant for the moderate severity levels of harm. Another study that is performed on a relatively smaller size of data showed similar findings [7].

The deviations may be caused by several reasons. Firstly, the reporters vary in background. Reporting is open to clinicians in the hospitals, including physicians, nurses, pharmacists, etc. Nurses and pharmacists are reported to be more active in the reporting because they witness errors more frequently during the course of care [9]. For example, they have more chances to witness and report medication errors. When they do, they are likely to assign medications errors with a lower harm score

compared to other clinicians [10,11]. Secondly, reporters' understanding of the harm scale exerts an influence on the rating [10]. Studies have suggested an important role of education and training in the reporting [12]. Besides, biased harm scores can be a result of unclear guidelines and information representation, such as the definitions [7] and knowledge structure [13]. Thirdly, the way in which events are reported may influence the reliability of harm scores. This argument is mainly centered on the capability of hospitals to discover and make adjustments of potentially biased scores. Compared to paper-based reporting, web-based reporting holds potential to disclose biased harm scores, as it is advanced in viewing aggregate data and trends.

Predicting Harm Scores from Patient Safety Reports

An alternative of calibrating biased harm score is to develop a mechanism of predicting harm score from patient safety reports. During the reporting, the decision of assigning a harm score to an event is made by reporters' understanding of the event, their experience, and perception of environment. While components such as experience and perception are subjective, the event itself is relatively more objective. As such, decisions that are purely based on the events are likely to reduce the bias caused by human. In most of the hospital reporting systems, events are storytelling-like and recorded in a text format, namely patient safety reports. These reports contain substantial and essential information to make judgment of harm scores. Most importantly, informatics techniques are available to extract information without human biases. Text classification is a candidate technique that purports to predict classes of text based on statistical regulations of term distribution in the text. To perform text classification, a statistical model is trained through learning term frequency from a set of categorized documents. The trained model is then capable of predicting uncategorized homogeneous documents with correct classes. This method has been broadly used in biomedical domain to reduce manual production time [14].

We propose that text classification can be used to predict harm scores based on patient safety reports. In this study, we will train classifiers from a set of reports that are assigned with harm scores according to the Common Formats Harm Scale (v1.2). The classifiers will predict harms scores of unlabeled reports where the performance of classifiers will be evaluated. From a practical perspective, the classification results are expected to eliminate potentially biased harm scores based on stored reports.

Table 1 – Distribution of harm scores among 2919 reports.

Harm Score	Meaning	Frequency
a	Death	11
b	Severe harm	36
c	Moderate harm	144
d	Mild harm	336
e	No harm	626
f	Unknown	1766

Methods

Data

The dataset consists of a corpus of 2919 de-identified patient safety reports from a university health care system. The reports cover a range of incident types that are labeled by reporters (see Figure 2).

The reports have been cross-validated by a group of domain experts, assigned with harm scores using the Common Formats

Harm Scale (see Table 1). In the text classification task, the assigned harm scores serve as the gold standard to be compared with machine prediction.

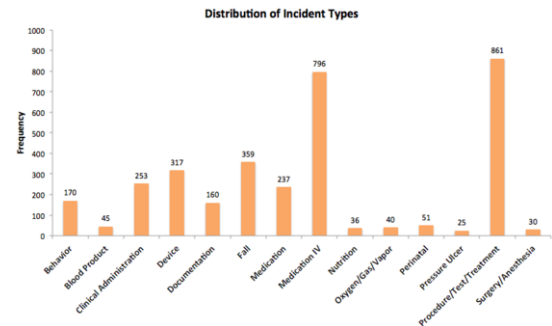


Figure 2 – Distribution of incident types.

Procedure

Environment

We provided a 64-bit OS system with a processing power of 2.2 GHz with 4 Cores 8 Threads and a memory of 8 GB RAM to perform the experiments. The classification experiment was performed on WEKA 3.6 [15].

Text Processing

To extract term frequency information from the raw data, we followed a serial of procedures to prepare the data. (1) Snowball stemmer was used to reduce inflected terms to their root form [16]. (2) Rainbow list was used to remove stop words [17]. (3) Alphabetic tokenizer was used to break a string of text into terms. (4) Lower case token was applied to all the terms. (5) TF-IDF (term frequency-inverse document frequency) was used in the transformation of documents into a bag-of-words (BOW) matrix keeping 1000 unique terms [18].

Text Classification Algorithms

We selected three types of algorithms that are well documented in text processing and biomedical application. They include decision tree algorithm, lazy algorithm, probabilistic algorithm, and support vector machine (SVM) [19]. For the decision tree, we employed C4.5 since it is reported effective in processing text [20]. For the lazy algorithm, we employed k-Nearest Neighbor (kNN) [21] for its well-balanced efficiency and predictive performance in medical text [22]. For the probabilistic algorithm, we employed Naïve Bayesian [23]. See Table 2 for a list of algorithms we used. A benchmark comparison is performed among these algorithms.

Table 2 – A list of selected algorithms.

Algorithm	Implementation	Parameter
Decision Tree	C4.5	
k-Nearest Neighbor	IBk	k = 1
Naïve Bayes	NaiveBayes	
Support Vector Machine	LibSVM	Linear SVM

Evaluation

For all the six harm scores and four algorithms, we employed a 10-fold cross validation to compare between algorithm performance. Each round of evaluation ran 10 times, which produced a total of 2,400 results. Performance was measured by F measure, which is a weighted average between precision and recall. The generic F measure is given as

F Measure

$$= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1)$$

In a classification task, precision is the fraction of retrieved documents that are relevant to a given label. It measures the ability of a classifier not to label a document as relevant when it is not.

$$\text{Precision} = \frac{y \cap \hat{y}}{\hat{y}} \quad (2)$$

where \hat{Y} denotes the predicted set of labels; Y denotes the exact set of labels.

Recall is the fraction of relevant documents that are retrieved. It measures the ability of a classifier to retrieve as more relevant documents as possible.

$$\text{Recall} = \frac{y \cap \hat{y}}{y} \quad (3)$$

In addition, we provided an estimate of receiver operating characteristic curve (ROC) as a metric for assessing the trade-off between true positive and false positive.

Results

Table 3 shows a ranking between the six tasks of classifying each harm score. The numbers indicate the number of wins or losses (negative number) of any task against the other tasks. On the metrics of precision, recall, and F measure, the task becomes more difficult if the harm score becomes smaller. This is probably because the narratives in the mild-harm or unknown-harm events contain less significant term frequency information that distinguishes the events from severe events. However, the ROC shows that tasks of classifying score 0 (unknown) and score d (mild harm) outperform others, indicating a well-controlled false positive.

Table 3 – Ranking test for classification tasks. ($p < .05$)

	Precision	Recall	F measure	ROC
f	-20	-18	-20	11
e	-12	-12	-12	-7
d	-4	-1	-2	11
c	4	0	2	0
b	13	14	12	-7
a	19	17	20	-8

Table 4 shows the results of ranking test for different algorithms. Naïve Bayesian outperformed in the ranking of precision and ROC. C4.5 is ranked the best algorithm on Recall and F measure.

Table 4 – Ranking test for different algorithms. ($p < .05$)

	Precision	Recall	F measure	ROC
C4.5	1	6	8	-3
kNN	-9	6	4	-9
Naïve Bayesian	9	-11	-10	18
SVM	-1	-1	-2	-6

Figure 3 shows the benchmark comparison on precision. Naïve Bayesian (precision = 0.88) and C4.5 (precision = 0.88) outperformed kNN (precision = 0.85) and SVM (precision = 0.87) in the comparison across all the six classification tasks. Paired t test shows that Naïve Bayesian performed better than C4.5 on tasks of classifying score e, d, c, and b, respectively ($p < .05$). But for the tasks of classifying score f and a, C4.5 performed better ($p < .05$).

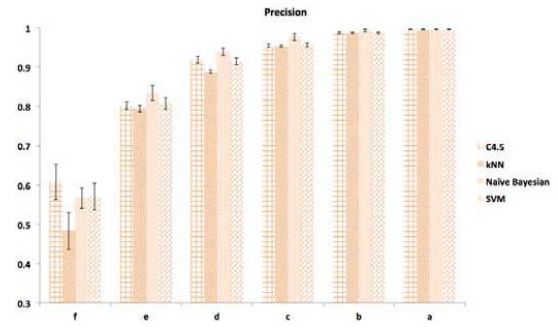


Figure 3 – Precision of algorithms by different harm scores.

Figure 4 shows the benchmark comparison on recall. C4.5 (recall = 0.90) ranks the best algorithm compared to kNN (recall = 0.88), Naïve Bayesian (recall = 0.83), and SVM (recall = 0.87).

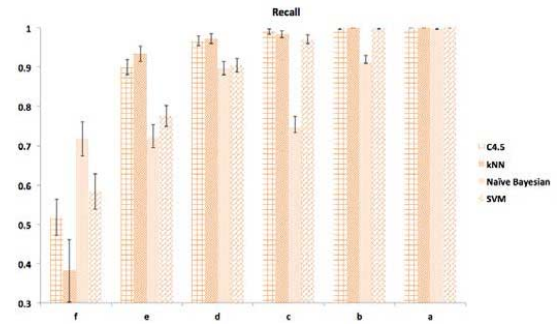


Figure 4 – Recall of algorithms by different harm scores.

Figure 5 shows the benchmark comparison on F measure. C4.5 ($F = 0.89$) ranks the best algorithm against kNN ($F = 0.86$), Naïve Bayesian ($F = 0.85$), and SVM ($F = 0.87$) overall.

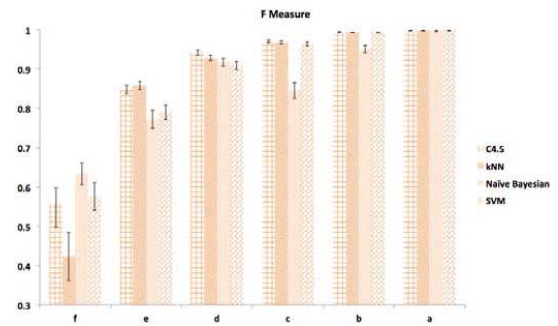


Figure 5 – F measure of algorithms by different harm scores.

Figure 6 shows the benchmark comparison on the area under ROC curve. Naïve Bayesian outperformed C4.5 (ROC = 0.76), kNN (ROC = 0.59), and SVM (ROC = 0.56) at all the classification tasks ($p < .05$).

Concerning the efficiency of algorithm, kNN used an average of 0.15 seconds of model training time, recognized the most efficient algorithm ($p < .05$) compared to C4.5 (time = 22.48 seconds), Naïve Bayesian (time = 0.92 seconds), and SVM (time = 1.70 seconds). An interaction of efficiency and performance is observed on Recall and F measure only, indicating that more time is needed for better performed algorithms (see Figure 7).

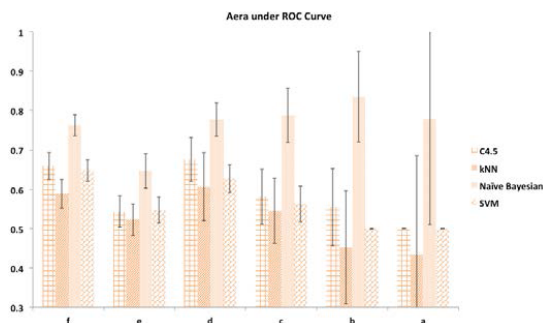


Figure 6 – Area under ROC curve of algorithms by different harm scores.

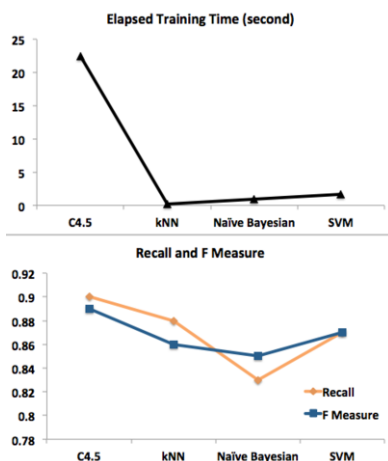


Figure 7 – Interaction between model training time and, recall and F measure, respectively.

Discussion

Experimental Findings

Our findings demonstrate the effectiveness of applying text classification to identify harm scores. The narratives of patient safety reports contain sizable information to determine the harm scores of events. In the experiment, the four types of text classification were well performed. However, the classification performance varied by the tasks of identifying different harm scores. There may be two factors that influenced the results. Firstly, the amount of reports of score a and b is significantly smaller, which might cause a label imbalance problem. This problem states that a classifier may be under trained because there are not sufficient reports in the minority labels. As an example, the classification tasks of identifying score a and b show outstanding performances on precision and recall but much worse ones on ROC. Secondly, levels of difficulties may vary by tasks. The task of identifying score f (unknown harm) is worst performed in terms of precision, recall, and F measure, but not ROC. Intuitively, the term frequency information in the reports of score f is sparse compared to the others.

Predicting harm scores based on patient safety reports is significantly efficient. Our findings confirmed a small computational cost of building the four types of classifiers but revealed some differences between these classifiers. Rule-based classifiers, i.e., C4.5 in our study, demand more time and computational

resources. On the contrary, lazy classifiers and probabilistic classifiers use comparatively less time and computational resources, indicating a much improved efficiency. However, there was a tradeoff between efficiency and predictive power. Although, C4.5, for example, is most time consuming in our experiment, it showed best performance on Recall and F measure. This effect implies that C4.5 has the best capability to identify as more true positive reports as possible, regardless of identifying false positive ones. Moreover, this capacity is still dominant on the comparison of F measure, which is a combined metric of precision and recall.

Selecting a suitable classifier is task dependent. If the task concerns more about false alarm, i.e. mistakenly assigning a report to an irrelevant harm score, C4.5 and Naïve Bayesian are better. If the task concerns more about retrieving more reports that belong to a given harm score, C4.5 is preferred. When it takes both factors into account, C 4.5 is recommended. Because it still won on the F measure by showing a statistically significant difference. In the practice, however, it is mostly concerned to enlarge the true positive rate and reduce the false positive rate. Thus, Naïve Bayesian is the best classifier.

Clinical Implementation

We envision that the automated classification of harm scores could assist in calibrating any human biases raised in the process of reporting. The proposed methods are implementable to most of the existing web-based reporting systems. In our view, text classification takes advantages of the existing reporting systems from three aspects.

Firstly, text classification corresponding to the narrative data that are commonly used in the reporting. Narrative data makes patient safety reports different from many other clinical data. A patient safety event is mostly encoded in a story-telling fashion since the elements of health care are complex and temporally organized. It is less likely to include many detailed and critical information in structured data, such as numeric or categorical data. This fact hinders many analytical methods that are applicable for generic clinical data.

Secondly, text classification promotes aggregate analysis and reporting. The Patient Safety and Quality Improvement Act of 2005 has called on to build a national mechanism of error reporting and analysis. Over the past decade, the rapidly increased volume of reported data has shown a quantitative improvement as well as a technique bottleneck of timely processing such a huge amount of data. Without a feasible solution, the analysis at a national level is of less practical value. The efficiency of our method suggests a practical use to be implemented in clinic, which is promising in largely reducing the demand of human labor.

Thirdly, the automated classification of harm scores is controllable because it is semi-supervised in practice. Caution should be taken when we apply automated method to medical decisions. We noted that all the evaluated classifiers are at a certain rate of error. Though this is not unique in our case, such an error should be controlled at a reasonable level. It is highly recommended to perform the classification under human supervision, as it is the case for most of the informatics tools implemented in medicine.

Limitation and Future Direction

We highlighted a need for creating an objective mechanism of overseeing human biases of categorizing harm scores. However, our study is of less value without a discussion of limitations. One limitation is the relatively small sample size used for the experiment. This problem may cause imbalanced labels,

which further harm the performance. In addition, a small sample limits the possibility to evaluate rare cases that may hold important clinical value but have limited distribution. Therefore, we will enrich the sample size and include a broader and more representative dataset.

The other limitation includes less consideration of label correlation, which may cast a crucial influence on the classification performance. The label correlation may not only be limited within the different harm scores but also with a number of categories such as contributing factors, settings, and procedures. In the future, we will evaluate such relational information by experimenting a multi-label classification.

Last but not least, the present study did not consider categorizing harm scores on temporal information. For example, the Common Formats Harm Scale (v1.2) consists of a two-point scale assessing the duration of harm. The extraction of temporal information from medical (e.g., clinical notes) has not been adequately investigated but has been on our research agenda.

Conclusion

This study aims to apply automated text classification to categorize harm scores from patient safety reports. This method is applicable to discover and calibrate potential classification biases caused by human judgment. Four types of classification were evaluated in the experiment, yielding effectiveness and efficiency of the proposed approach. In addition, this approach holds promise in facilitating the labor intense analysis of large volume patient safety reports which is in accordance with the goal of establishing a nationwide safety reporting mechanism.

Acknowledgements

The study is supported by a research grant (1R01HS022895) from the Agency of Healthcare Research and Quality, and University of Texas System Grants Program (#156374).

References

- [1] P.J. Pronovost, L.L. Morlock, J.B. Sexton, M.R. Miller, C.G. Holzmueller, D.A. Thompson, L.H. Lubomski, and A.W. Wu, Improving the value of patient safety reporting systems, *Advances in patient safety: new directions and alternative approaches* **1** (2008).
- [2] L.T. Kohn, J.M. Corrigan, and M.S. Donaldson, *To Err Is Human: Building a Safer Health System*, National Academies Press, 2000.
- [3] C.M. Clancy, Common formats allow uniform collection and reporting of patient safety data by patient safety organizations, *American Journal of Medical Quality* **25** (2010), 73-75.
- [4] H. Sherman, G. Castro, M. Fletcher, M. Hatlie, P. Hibbert, R. Jakob, R. Koss, P. Lewalle, J. Loeb, T. Perneger, W. Runciman, R. Thomson, T. Van Der Schaaf, and M. Virtanen, Towards an International Classification for Patient Safety: The conceptual framework, *International Journal for Quality in Health Care* **21** (2009), 2-8.
- [5] D.C. Classen, R. Resar, F. Griffin, F. Federico, T. Frankel, N. Kimmel, J.C. Whittington, A. Frankel, A. Seger, and B.C. James, 'Global trigger tool' shows that adverse events in hospitals may be ten times greater than previously measured, *Health Affairs* **30** (2011), 581-589.
- [6] M. Hoppes and J. Mitchell, Serious Safety Events: A Focus on Harm Classification: Deviation in Care as Link, 2014.
- [7] T. Abbasi, D. Adornetto-Garcia, P.A. Johnston, J.H. Segovia, and B. Summers, Accuracy of Harm Scores Entered Into an Event Reporting System, *Journal of Nursing Administration* **45** (2015), 218-225.
- [8] T. Williams, M. Szekendi, S. Pavkovic, W. Clevenger, and J. Cereze, The reliability of AHRQ common format harm scales in rating patient safety events, *Journal of patient safety* **11** (2015), 52-59.
- [9] S.M. Evans, J.G. Berry, B.J. Smith, A. Esterman, P. Selim, J. O'Shaughnessy, and M. DeWit, Attitudes and barriers to incident reporting: a collaborative hospital study., *Quality & safety in health care* **15** (2006), 39-43.
- [10] S.D. Williams and D.M. Ashcroft, Medication errors: how reliable are the severity ratings reported to the national reporting and learning system?, *International Journal for Quality in Health Care* **21** (2009), 316-320.
- [11] P.J. Pronovost, C.G. Holzmueller, J. Young, P. Whitney, A.W. Wu, D.A. Thompson, L.H. Lubomski, and L.L. Morlock, Using incident reporting to improve patient safety: a conceptual model, *Journal of patient safety* **3** (2007), 27-33.
- [12] A.M. Mayo and D. Duncan, Nurse perceptions of medication errors: what we need to know for patient safety, *Journal of nursing care quality* **19** (2004), 209-217.
- [13] C. Liang and Y. Gong, Knowledge Representation in Patient Safety Reporting: An Ontological Approach, *Journal of Data and Information Science* **1** (2016), 75-91.
- [14] A.M. Cohen and W.R. Hersh, A survey of current work in biomedical text mining, *Briefings in bioinformatics* **6** (2005), 57-71.
- [15] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, The WEKA data mining software: an update, *ACM SIGKDD explorations newsletter* **11** (2009), 10-18.
- [16] M.F. Porter, Snowball: A language for stemming algorithms, 2001.
- [17] A. McCallum, Rainbow project, 1998.
- [18] G. Salton and C. Buckley, Term-weighting approaches in automatic text retrieval, *Information processing & management* **24** (1988), 513-523.
- [19] C.C. Chang and C.J. Lin, LIBSVM: a library for support vector machines, in: *ACM Transactions on Intelligent Systems and Technology (TIST)*, ACM, 2011, p. 27.
- [20] J.R. Quinlan, *C4. 5: programs for machine learning*, Elsevier, 2014.
- [21] J. Wang and J.D. Zucker, Solving multiple-instance problem: A lazy learning approach, in: *Proc. 17th International Conf. on Machine Learning*, Morgan Kaufmann, 2000.
- [22] C. Liang and Y. Gong, Enhancing Patient Safety Event Reporting by K-nearest Neighbor Classifier. Context Sensitive Health Informatics, in: *Context Sensitive Health Informatics*, 2015.
- [23] G.H. John and P. Langley, Estimating continuous distributions in Bayesian classifiers, in: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 338-345.

Address for correspondence

Yang Gong, MD, PhD
7000 Fannin St. Suite 600, Houston 77030 Texas
Yang.Gong@uth.tmc.edu
Tel: +1 713 500 3547

The Value of Data and the Value of Questions: Achieving Improvement in Hospital Pharmacy and Medication Safety

Valentina Lichtner^a, Tony Cornford^b

^a Leeds University Business School, University of Leeds, Leeds, UK,

^b Department of Management, London School of Economics, London, UK

Abstract

Medication safety incidents are one of the major patient safety issue faced across all healthcare services and one that is very challenging to tackle. To make progress, data about the supply and use of medicines that is generated and made available in clinical systems can serve both the purposes of patient safety and service quality improvement. This paper shows how the value of data for these purposes can be framed in terms of the value of questions. This theme is developed based on an interview with a quality and safety pharmacist working in a large hospital pharmacy unit.

Keywords:

Task Performance and Analysis; Pharmaceutical Services; Patient Safety

Introduction

“Big Data is not about providing answers to questions we could not answer before; it is about our ability to think of the new questions” [1].

This is the kind of advice the world of business consulting gives to companies wishing to generate value from Big Data. The advice is to focus on questions [1; 2], and especially new ones. This paper explores how this argument applies to the running of hospital pharmacy services. It presents a case study, a pharmacy perspective on quality improvement in medication safety, in the context of an implementation of an electronic prescribing and administration (EPMA) system and the new data it can provide. The case is centred on a conversation with a practitioner. We asked about the value of data, and she reminded us of the value of questions.

There is an extensive body of knowledge on the use of big data in the healthcare domain; data about patients and patient care activities are increasingly researched for clinical purposes (e.g. predictive analytics) and commissioning. In relation to medicines, big data and associated algorithms are offering insight for pharmacogenomics and precision medicine [3; 4]. The recent seminal paper by Google Research team [5] also shows how machine learning through hospital based data (on eye images) can produce diagnostic systems, specifically, in this case, for detecting diabetic retinopathy. Big data research is more limited on how big data are used by practitioners to manage healthcare services, from a business operations perspective, with some evidence emerging from the trade press [6; 7]. Visual analytics, for example in the form of dashboards, are proposed as one way for service managers of dealing with data complexity, for example for purposes of quality and safety [8; 9].

In this paper the main concern is data about medicines for purposes of improving quality and safety in a hospital setting. Medicines are the most used therapeutic intervention and one

frequently associated with preventable adverse events. Medication safety is one of the major patient safety issue faced across all healthcare domains and very challenging to achieve.

Methods

The case presented in this paper is taken from data collected for a project investigating the digitalisation of supply and use of medicines in a hospital setting. Data collection included non-participant observations, interviews and documentary analysis. Observations of staff (prescribers and dispensers) work practice were carried out to gather data on the use of data and technology for the supply of medicines in these settings. Semi-structured interviews were conducted with a variety of stakeholders including patients, doctors, nurses, pharmacists and managers. Interviews explored views and experiences with new technologies and services aimed at supporting supply and use of medicines. With participants' consent, interviews were recorded and transcribed.

Overall data were collected through 37 interviews and about 103 hours of observation. This paper focuses on one interview transcript, with a pharmacist, manager of the service responsible for safety and quality improvement in the hospital pharmacy. This interview was chosen for its relevance and for the insight it offers into uses and challenges of data for hospital management. All quotes cited in the Results section of this paper come from this interview transcript.

Results

The setting

The site we studied is a large National Health Service (NHS) teaching hospital, covering five buildings located in different parts of one of England's largest cities. Medicines used in the hospital are managed by the Pharmacy Services Unit, through a number of on-site dispensaries and about 500 professional staff. The Unit also provides specialized pharmacy services for the local and nearby communities. The Unit comprises a number of pharmacy 'service areas', including procurement, supply, information services and pharmacy IT, and quality and safety.

From a service perspective, the Pharmacy Services Unit is running a business, servicing other units in and beyond the hospital, such as the clinical wards and local community clinics – these representing 'clients' and 'users' of the services. In the NHS the cost of medicines is covered by budgets held locally by Clinical Commissioning Groups (CCGs), ultimately responding to the Department of Health; the local CCG is thus also a 'client' of the Unit, the one paying for the service.

Pharmacy services' activities are run and managed through a number of digital information systems, including; a supply

"...the patient wants to know how quickly can [their medicine] be with them, ready for them to take it home, and we know that there's a bit of a delay between it being ready in the dispensary and it then getting to the ward, and so you could argue that one of our measures ought to be: when it's ready in the dispensary how long does it [then] take to get to the ward?"

Gathering the data

The data for performance assessment, and the layers of data displayed in dashboards, come from a number of systems – some digital, other requiring purpose-built auditing tools which are mainly paper based. The processing of digital data and especially data collection by hand, are resource intensive and time consuming, and not to be lightly introduced.

"...they are different [pharmacy units], and there is no one system of data collection that we use across all ..."

"[data] come from a collection of systems and a collection of manually collected, or data that's taken from other systems and then amended or extracted for our own purposes. So [...] currently it's a very [pause] I was going to say it's very manually heavy. It's, it's less so now because we've worked really hard to only use electronically-collected data, but there are still some bits of data we have to collect by hand."

At the time of the study the hospital was implementing an EPMA system integrated into a patient record portal and the expectation was that these integrated systems would be able to provide data, without the need for manually auditing paper charts or using data collection forms.

"... when every patient has their drug chart in [EPMA], then that will sit within [the EPR] and the data will be shared and you will just be able to extract it in one transaction."

The availability of the right data needs planning

When data need to be collected manually, you have to plan ahead for that data collection; thus have to know your question. Once data are digital, at least in theory, there is a potential to answer emerging questions that were not planned for. However, in practice there may be a digital system in place but the data may not have been entered appropriately, or the system may not actually record the data you were expecting to have. Thus you still have to plan ahead, and 'be clear what you are going to measure'.

"Well I think it depends because you can only extract the data that you put in, and so if your system is not collecting the data that you require, then you still have to plan."

"... so a good example is: on our [paper] drug chart we have something called a Day 3 Antimicrobial Review section, so at the moment if you do a Day 3 you tick that box [on paper] and put a code in it, and we go along and [manually audit the charts] we measure the number of patients who've had a Day 3 review. Interestingly in [EPMA], there is no such box to tick and so the question is: how would you extract the data from [EPMA]? What data is recorded to show that a Day 3 review has occurred?"

"...., you still have to plan and be clear what you're trying to measure, because just saying 'it will be there', isn't always the case."

Some data are still imperfect answers to your questions

Despite planning and efforts of data collection, the data that can be available and collected, are not necessarily matched to the question of interest. The case of monitoring for antimicrobial use in hospital is again one such example:

"Well you can go on and there is a box for each ward to use which says, 'Is the patient on antimicrobials?' and you can ac-

tually manually put it in. [...] It just says they're on antimicrobials, 'yes, no'. So we're working with imperfect data because the data we can often get doesn't answer the question we want answering ..."

The question must be worth the effort

Furthermore, not all questions are worth the data collection efforts (especially the manual audits); the overall aim is achievement of improvement and this gives a measure of the value of the data (and the value of the questions).

"... we were using our pharmacy teams to collect huge amounts of manually-collected data [...] and we felt that actually we were spending a lot of resource in collecting data that was disproportionate to the improvements we were looking to achieve."

From operations' performance to medicines' questions

Beyond the performance of the Pharmacy Unit, achievement of overall business objectives is the safe and appropriate use of medicines – the right medicines given to the right patients at the right times (the 'three rights' that are the goal of most pharmacy services' activities). The hope is that digital data about medicines will enable answering questions about the right (or wrong) supply and use of medicines. These will be expressed in increasingly complex questions – e.g. not just about actions but about rationales for actions. The hope is also that answers to these types of questions will be able to inform different ways of improving quality and safety and medicines use.

Questions about the use of antibiotics are an example because they encompass both the safety of individual patients, the outcomes of patients treated with that medicine (e.g. for reasons of research), and also more generally stewardship to tackle the challenges of antimicrobial resistance [10] and achieve wider societal benefits. Using antibiotics (antimicrobials) as an example, the extracts below show the connection between the different steps in the process, from one question, to data and then more questions, and eventually identification of areas for improvement (such as providing training or better information for doctors).

"So antimicrobials is a really good example and [...] [we have] very specialist people who are very interested in the fine detail of which drug by which consultant to which patient, ..."

"..., but then there is another element [of interest, about patient safety] of which patients: has anybody tried to prescribe a drug, an antimicrobial say, to a patient who's classed as allergic to an antimicrobial?"

"... because there are some patients who the risk of giving a medicine to, is much less than risking them having an allergy, but then is equally the one in a million where it's a serious life consequence. [...] our job is to identify the patients it's okay to do it in and make sure that there's a good rationale for doing it, and the ones that it's not okay to do it in..."

"... [with the data from EPMA] we will have, if you like, the ones who have tried to prescribe it and continue to prescribe it - and then we'll have for those a reason for them continuing to prescribe it - and we'll also be able then to see how many doses the patient had and the outcome of the patient. ..."

"We'll also be able to see the number of doctors who tried to prescribe [for] a patient who was allergic a drug and then chose not to proceed. And they are, they're equally of interest to me because actually they tell us something about whether the system [...] has introduced a barrier to improve safety or not."

"So what we would then need to know is: were those doctors or prescribers trying to prescribe something because they had lack of information or just because they didn't think about it, or what

was the rationale, and at the moment the only thing we can explore is, is where it's gone wrong, ..."

From (retrospective) questions to (prospective) guidance to use

The meta-question posed is 'what are we really trying to achieve' with the data and by answering those questions. The overall aim of knowing whether patients have been given the right (or wrong) medicine is to make sure that it does (or does not) happen again, and that it does (or does not) happen to other patients. The digital system may enable a shift from retrospective data collection and use, to prospective 'decision support' and behavioural change. The idea is: instead of, or as well as, trying to collect digitally the same data that were collected on paper (such as a 3 day antibiotic review), change the data points to generate different behavior at the point of care.

"... actually you could [...] say, "Well you might change your measure from Day 3 to a 48 hour review", so if the system has something that collects a 48 hour review in it, you would ask yourself, "Well do I really need a Day 3 box ticked?" So yes, there is a bit of what are you trying to achieve with that number, and actually if you go back to the patient experience and the patient safety bit, we're trying to make sure that only patients who need antibiotics get them, and part of the reason for the Day 3 is that's the point at which you should have all of the information to make a decision. Well, you might be able to, if our systems were slick, [...] make a decision at 48 hours, which would be even better because that would mean that you're not exposing patients to a longer treatment than they need to have of a certain antibiotic, and so it might help us smooth that out, absolutely, and I'm a great believer in once you can scan data you might say, "Well actually I'd rather have that data than this data," and that data will help me answer a question differently. So I absolutely think it will do that..."

The business of the pharmacy meeting 'clients' and patients' needs

Returning to the service delivery perspective, data is only a means to an end, and the overall goal is client satisfaction - mainly patients, but also clinicians using the services of the hospital pharmacy, and the clinical commissioning groups paying for them. Satisfaction and user experience are better assessed through gathering qualitative data and user feedback. It is not all about data points, but just as much about relationships, communication and trust. And it is also about being able to answer or pre-empt *their* (commissioners', patients, clinicians) questions.

... for me, data is a single tool that we use to ensure quality and safety, so my bigger remit around governance quality and safety is about making sure that we do everything that we should be doing, that the outcomes are the outcomes we are looking for and that our patients are looking for, and so data is only one element of that because data is only about numbers, not about experience, and you could have the best numbers in the world and have a poor experience.

Because if the delivery of the service is spot on but doesn't have your patient as the focus, then you're not getting the outcome.

So there's a bit of qualitative data and a bit of quantitative data that you need to look at in tandem.

[...] So patient, direct patient feedback about their experience is probably the gold standard, but also not forgetting that what we do is also with our staff, so nursing staff's feedback about the availability of medicines in their workplace. Feedback from doctors about their training on prescribing or the availability

of information to help them prescribe safely, that type of information about our service providers is also really important to us.

"And our commissioners... [...], because our commissioners commission us to use medicines in certain ways and so some of what we monitor is to make sure that as an organisation we're only using those medicines to treat the patients that we're asked to treat, [...]. So there are some very expensive medicines usually, or medicines with a poor safety profile, and collectively we might say, "Well we only want to use [...] that drug to treat those patients with a specific condition or a specific set of circumstances", and part of our job is to make sure that that drug is only ever used for those patients. [...]"

"So [for example] there are some medicines that we are told by commissioners that they only want us to use in patients with cystic fibrosis who have a very low respiratory measure and what we have to ensure is that we only ever prescribe it to those, that group of patients. It's a bit like a NICE guideline, you know, NICE guidelines will say, "Only use this drug in these circumstances", and that's where eMeds will come into its own, because we will be able to say, "Tell me all the patients who've got that drug and tell me if they [meet] that criteria", and then we'll be able to say, "Yes that drug is only being used in patients with that criteria" or "There's ten patients who you used it in who didn't meet those criteria. Let's go and have a look at the circumstances of that". So, but actually part of our assessment is, "Do our commissioners trust us that we're doing what they're asking us to do?" and some of that's around data but some of that's also around how we interact with our commissioners."

Discussion

In this case study of a pharmacy service unit, *concerns about data* are framed in terms of *concerns for questions*: what are the questions to ask and what questions data might be able to answer.

Digitalization of clinical care can certainly make data collection at the point of care less resource intensive and less time consuming, potentially making some of it available in real time. However, it also inevitably changes the type of data that is gathered, the opportunity costs implied. The data that were available in the past may not be available through the new digital systems, and traditional collection methods may need to be still used.

Overall, this may change the questions that can be asked and answered, and leads to a rethink of what the original questions were for, whether new questions would better achieve the overall aim. In particular new data ecologies can offer the potential for more holistic and overarching types of question relating to customer service, rather than task performance. Furthermore, the act of gathering data at the point of care (through clinician's data entry into digital systems or by embedded technologies) for purposes of retrospective data analysis, can be transformed into information points for prospective guidance for action – such as when asking about patients' allergies to antibiotics turns into a change in prescribers' choice of medicine to treat the patient's bacterial infection.

Limitations

This study is limited to only one interview, in one hospital, in one specific context, and should not be taken to represent the business of hospital pharmacies across all NHS hospitals in England or beyond. The interview is however taken from a larger dataset, and informed by observations of pharmacy and clinical activities across different hospital areas.

Conclusions

Digital data gives us the possibility of answering new questions, changing the questions we ask, and making questions and data-driven responses a more active part of health care delivery. It is by posing the right question that improvement in quality and safety can be achieved - not by access data alone.

Acknowledgements

The Delivering Digital Drugs study is funded by Research Councils UK / 'New Economic Models in the Digital Economy' Programme (EP/L021188/1). NHS Research Ethics Committee reviewed and granted approval for the study (REC reference 15-YH-0568).

References

- [1] F. Wammes, Big Data: It's the Question to Solve, Not the Answer to Give, in, CapGemini, 2015.
- [2] B. Marr, Big Data: Too Many Answers, Not Enough Questions, in, Forbes, 2015.
- [3] K.H. Yu, S.N. Hart, R. Goldfeder, Q.C. Zhang, S.C. Parker, and M. Snyder, Harnessing big data for precision medicine: infrastructures and applications, *Pac Symp Biocomput* **22** (2016), 635-639.
- [4] J.C. Denny, Surveying Recent Themes in Translational Bioinformatics: Big Data in EHRs, Omics for Drugs, and Personal Genomics, *Yearb Med Inform* **9** (2014), 199-205.
- [5] V. Gulshan, L. Peng, M. Coram, M.C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P.C. Nelson, J.L. Mega, and D.R. Webster, Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, *JAMA* **316** (2016), 2402-2410.
- [6] S. Hoppszallern, MOST WIRED. Creating health care value with big data intelligence, *Hosp Health Netw* **90** (2016), 20.
- [7] S.S. Desai, J. Wilkerson, and T. Roberts, Improving the Success of Strategic Management Using Big Data, *World Hosp Health Serv* **52** (2016), 49-51.
- [8] J.J. Caban and D. Gotz, Visual analytics in healthcare--opportunities and research challenges, *J Am Med Inform Assoc* **22** (2015), 260-262.
- [9] R.M. Ratwani and A. Fong, 'Connecting the dots': leveraging visual analytics to make sense of patient safety event reports, *J Am Med Inform Assoc* **22** (2015), 312-317.
- [10] M.T. Baysari, E.C. Lehnborn, L. Li, A. Hargreaves, R.O. Day, and J.I. Westbrook, The effectiveness of information technology to improve antimicrobial prescribing in hospitals: A systematic review and meta-analysis, *Int J Med Inform* **92** (2016), 15-34.

Address for correspondence

Valentina Lichtner, v.lichtner@leeds.ac.uk

User-Centered Design Improves the Usability of Drug-Drug Interaction Alerts: A Validation Study in the Real Scenario

Daniel R Luna^{a,b}, Daniel A Rizzato Lede^a, Luciana Rubin^a, Carlos M Otero^a, Juan M Ortiz^a,
Mónica G García^a, Romina P Rapisarda^a, Marcelo R Risk^{a,b,c}, Fernán González Bernaldo de Quirós^d

^a Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina,

^b Instituto Tecnológico de Buenos Aires (ITBA), Argentina,

^c Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina,

^d Medical Vice-Direction of Strategic Planning, Hospital Italiano de Buenos Aires, Argentina

Abstract

Decision support systems can alert physicians to the existence of drug interactions. The Hospital Italiano de Buenos Aires, Argentina, has an in-house electronic health record with computerized physician order entry and clinical decision support. It includes a drug-drug interaction alert system, initially developed under traditional engineering techniques. As we detected a high alert override rate, we rebuilt the knowledge database and redesigned the alert interface with User-Centered Design techniques. A laboratory crossover study using clinical vignettes showed that new alerts were more usable than traditional ones. This paper aimed to validate these results through a controlled and randomized experimental study with two branches (old vs. new design) in a real setting. We analyzed, quantitatively, every fired alert between April 2015 and September 2016. Finally, we performed user surveys and qualitative interviews to inquire about their satisfaction and perceptions. In real scenarios, user-centered design alerts were more usable, being more effective and satisfactory, but less efficient than traditional alerts. "Safe omission", as a new concept, emerged from our stratified analyses and interviews.

Keywords:

Drug Interactions; Expert Systems; Software Design

Introduction

In 1999 the Institute of Medicine released the report "To err is human", stating that a quarter of total medical errors were related to medication [1]. Even though most errors are harmless, some of them can cause variable damage, including death. Fifty percent of these errors take place during drug prescription [2]. A common mistake is to neglect drug-drug interactions (DDI) [3]. Clinicians awareness of DDI can prevent related adverse drug events, but at least half of the time they are not recognized [4]. Clinical decision support is "the use of information and communication technologies to bring relevant knowledge to bear on the healthcare and well-being of a patient" [5]. For example, if physicians enter medication orders electronically, these systems can show drug safety alerts, including overdoses, duplicate orders and drug-drug interactions. Therefore, clinical decision support systems seem to be an appropriate solution, as they have shown improvements in both quality of care and resource optimization [6–8]. Although, several studies showed their low performance and high override rate [9–12]. From our perspective, these systems have at least five potential drawbacks: Excessive alerts (mostly with low clinical significance) due to imprecise knowledge databases, leading to alert fatigue [13]; Low quality interfaces,

lacking of workflow integration and intuitive design [14]; Lack of context information in the system hinders complex rules [15,16]; Absence of alert monitoring prevents improvement processes [17]; and Implementation variability, even for the same vendor, secondary to standards deficit [18].

In the mid-2000s, the Hospital Italiano de Buenos Aires, Argentina, implemented an in-house electronic health record system with computerized physician order entry (CPOE). Shortly after, we launched a clinical decision support system for drug-drug interaction alerts, developed with traditional software engineering. Clinical pharmacology experts created a local knowledge database in Spanish. We monitored periodically the CDSS performance and found a high alert override rate. Thus, we first focused on improving the knowledge database quality [19]. Our analysis included the systematic evaluation of each DDI according to clinical relevance, to eliminate combinations with a low probability of harm (false positives), as suggested in a recent consensus [20]. We also adapted the Lexicomp® alert severity tiers according to potential reaction seriousness [21], following recommendations from Paterno et al. [13]. Only the two highest risk ratings (D and X) were considered clinically relevant and triggered intrusive alerts. As alert acceptance remained persistently low, the chief medical information officer ordered the DDI alert system withdrawal, to look for other potential issues. Previous research by Seidling et al. found that the alert display quality most strongly predicted DDI alert acceptance [22]. Therefore, we endeavored to improve the alert design through User-Centered Design (UCD) techniques, as they have demonstrated to increase adoption and usage efficiency of health information technology tools [23]. According to Patel and Kannampallil, human-computer interaction is a fundamental aspect to consider when developing computer systems [24]. UCD is a process framework that makes a system usable and understandable by accounting for end-users' needs, wants and constraints, through the whole product cycle. Following perspectives from Norman, UCD starts by understanding and specifying the context and requirement analysis, and then designing and iteratively testing solutions [25]. This systematic process is regulated by ISO 9241-210 "Human-centered design for interactive systems" [26]. For Kushniruk et al. participatory design goes beyond UCD and cooperative design approaches to include end users as active participants in the design and decision making [27]. We started the analysis with an heuristic evaluation [28] of the standard alert, noting several issues regarding minimalism, consistency, feedback, visibility and documentation. Then, we reformed the DDI alerts by using a participatory design approach. As described in a state-of-the-art reference handbook for the

subject, participatory design can be defined as “a process of investigating, understanding, reflecting upon, establishing, developing and supporting mutual learning between multiple participants in collective ‘reflection-in-action’; the participants typically undertake the two principal roles of users and designers where the designers strive to learn the realities of the users’ situation while the users strive to articulate their desired aims and learn appropriate technological means to obtain them” [29]. A team of three health informatics specialists and two usability experts worked with final users following the ISO 9241-210. This phase took place at the HIBA from September 2013 to April 2014, and was undertaken in three stages (inquiry, participatory design, and usability testing), as described in a previous publication [30]. The whole process was iterative; each stage included prototyping cycles for domain saturation to reach the best possible model. The participants were physicians that worked in outpatient and inpatient settings. Fictitious patient scenarios (clinical vignettes) were developed based on real clinical cases [31], taking the most frequent and significant examples of DDI from our clinical data repository [11]. The last UCD prototype was developed as a new software version [30]. Afterwards, we performed a laboratory crossover study to test its usability, using new clinical vignettes. We found that new alerts were more usable than traditional ones, regarding efficiency, effectiveness, and user satisfaction [32]. A deep insight of this study was published recently in the Journal of Biomedical Informatics [33].

This paper aims to validate the lab results in a real scenario, through a controlled and randomized experimental study, measuring the same variables as preceding studies (efficiency, effectiveness and user satisfaction).

Methods

Setting

The Hospital Italiano de Buenos Aires (HIBA) is a non-profit healthcare academic center founded in 1853, with more than 2,700 physicians, 2,700 other health team members (including 1,200 nurses), and 1,800 administrative and support employees. The HIBA network includes two hospitals in Buenos Aires city and its suburban area, 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics and 150 associated private practices. It has a Health Maintenance Organization (Plan de Salud) that covers more than 150,000 people and provides health services to another 1,500,000 people who are covered by affiliated insurers. Between 2013 and 2014, the HIBA admitted more than 45,000 inpatients, conducted 45,000 surgical procedures (50% ambulatory) and 3,000,000 outpatient visits. The HIBA is a teaching hospital, with more than 30 medical residency-training programs, 34 fellowship programs and 400 residents and fellows in training.

Since 1998, the HIBA has run an in-house developed health information system, which includes clinical and administrative data [10]. Its Electronic Health Record system called Itálica, is an integrated, modular, problem oriented and patient centered system that works in different clinical settings (outpatient, inpatient, emergency and home care). Itálica allows computerized physician order entry for medications and medical tests, and storage and retrieval of test results, including archived images. It was the first hospital in Argentina and the second in Latin America to be certified by the HIMSS as level 6+ in the Electronic Medical Record Adoption Model. In recent years, our Health Informatics Department at the HIBA prioritized UCD in the design and development culture to enhance the usability of healthcare software. We conducted lectures, launched a pilot project, and assembled a usability team for service and dissemination [34].

Methodological Design

The HIBA Institutional Review Board approved the research protocol. The study was performed in a tertiary academic center, with users from different settings (outpatient, critical, and non-critical inpatient). For our experimental study, we randomly assigned physicians (system users) to two branches. We compared two different DDI alert interfaces: the standard one (developed under traditional techniques) and the participatory design model, generated under UCD techniques. Drug-drug interaction alerts were reinstated to the prescription system in April 2015. The clinical decision support system used the same DDI knowledge database and inference engine from previous stages of the study. The alert system ran every time a new prescription was placed and searched the knowledge database for potential interactions between each substance already on the list and the new drug. When an interaction was detected, the system opened a DDI alert modal (a dialog box or pop-up window that was displayed on top of the current page). Depending on the assigned branch, the physician would see the standard or the UCD version of the alert. The standard DDI alert interface can be seen in Figure 1.



Figure 1— Standard DDI Alert Interface (letters correspond to descriptions): the two drugs (A and B), the clinical significance (C), a brief explanation (D), a “learn more” button (E), and the action buttons to cancel the prescription (F) or ignore the alert (G).

Compared to the standard version, the new interface version had a different communication message, changing the displayed elements, warning colors and proposed actions (as shown in Figure 2). New output actions were specifically created for the novel DDI alerts.

Study Population

Physicians fulfilling the exclusion and inclusion criteria were selected for each clinical setting (outpatient, non-critical, and critical inpatient). All had worked with CPOE but had not been recently exposed to the original DDI alert, because it had been withdrawn more than 3 years earlier. Inclusion criteria were as follows: working as a physician of any specialty in the hospital for more than a year by March 2015. Physicians with previous participation in alert design or test stages were excluded. Users were randomized using their hospital ID to one of the two branches: the original interface and the participatory design version. None of them were aware of this allocation.

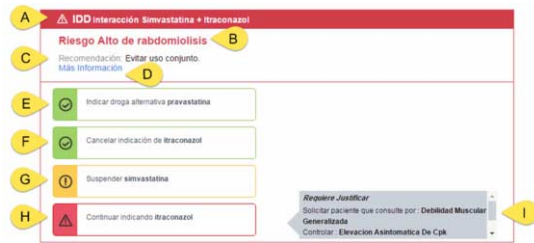


Figure 2– New Software Version Interface (letters correspond to descriptions) displaying the drugs involved and their risk (A), the drug interaction risk and severity (B); the main recommendation (C); a “learn more” link (D); the action buttons to “prescribe an alternative drug” (E), “cancel drug 2 prescription” (F), “cancel drug 1 prescription” (G) and the “keep both drugs” option to ignore the alert (H), which required justification and monitoring (I).

Measurements

We retrieved data from every fired alert from May 1st 2015 to September 30th 2016. We analyzed the metadata from each of them, including: the date and time, number of clicks on the “learn more” link and on the action buttons (“accept”, “cancel” and others), and the text for override justification. We also retrieved participants’ demographical information. All data was treated confidentially, and accessed only by authorized researchers. As in the previous lab study, efficiency, effectiveness and user satisfaction were selected as usability metrics to compare the performance of the new UCD alert against the standard version. Efficiency was measured as the time required to resolution. We defined that the alert was solved when the user finished the tasks and closed the modal window. As we could not witness participants’ reactions as in the lab study, we measured effectiveness according to the actions taken within the alert: it was “accepted” if the user decided not to prescribe the causative medication (or chose other alternative option), and was “canceled” if he kept the prescription. One way of identifying confounding is to examine the primary association of interest at different levels of a potential confounding factor. Therefore, we performed stratified analyses for both types of interfaces depending on different criteria: Interaction risk rating (severity); Setting: outpatient, critical, or non-critical inpatient; User seniority; Pair of drugs (combinations). Regarding user satisfaction, a brief survey translated and adapted from Zheng et al. [35] was conducted. It was similar to a Likert scale, in which the respondent indicated the degree of agreement with the satisfaction statements on a 4-point scale: totally disagree, disagree, agree, and totally agree. We automatically sent the questionnaires through SurveyMonkey® [36] to every user exposed to an alert. The message started with a short explanation on the subject, and showed a sample image of the alert that has been previously displayed. The system sent a reminder email after 48 hours. We closed the poll after 25 responses from each branch, considering domain saturation. We also collected user perceptions through direct qualitative interviews to a sample of exposed physicians. They were invited by email and we scheduled a meeting with those who accepted.

Statistical Analysis

Descriptive statistics were presented for all variables in the comparative study. Interval variables were parameterized by mean, median and quartiles. For categorical variables, the observed frequency (total number of observations within the category) and relative frequency percentages were used. Statistical analyses for all tests were performed using the R

software environment from R Project for Statistical Computing [37]. Statistical significance was considered when the probability was lower than 0.05.

Results

We analyzed the metadata of every triggered DDI alert from May 1st 2015 to September 30th 2016, within the hospital CPOE. There were 310 DDI alerts shown to different physicians: 168 (54%) were traditional alerts and 142 (46%) were UCD alerts. See Table 1 for demographical information. From 4141 drug interactions included in the knowledge database, the highest severity risk rating D and X represented 10% (440 combinations). In this study, there were just 94 pairs of drugs (2%) involved in the alerts during the yearly analysis.

Table 1– Demographics of Participants from both Branches. Year values are expressed as median (Q1 –Q3).

	Traditional Alert	UCD Alert	p
N	168 (54%)	142 (46%)	
Age (years)	32 (28-34)	31 (29-34)	0.92
Gender	F=57 %	F=56 %	
Seniority (years)	3 (2-4)	3 (1-4)	0.68

Time measured for efficiency was taken from the moment the window popped up to its closing. Table 2 shows that UCD alerts required less time for completion than traditional ones.

Table 2– Alert Resolution Efficiency, with time in seconds expressed as median (Q1 –Q3).

Variable	Traditional Alert	UCD Alert	p
Time (seconds)	17 (9-25.5)	10 (5-20.5)	0.009

Effectiveness was measured by the amount of accepted and canceled alerts. The global analysis is shown in Table 3.

Table 3– Alert Resolution Effectiveness, as absolute and relative (%) frequency of accepted and canceled alerts.

Result	Traditional Alert	UCD Alert	p
Accepted	84	42	
Canceled	84 (50%)	100 (70%)	<0.001

The stratified analysis regarding the risk rating of the interaction is shown in Table 4. Regarding setting (outpatient, critical, and non-critical inpatient), seniority, and specific drug combinations stratified analyses, there were no significant differences between both branches.

Table 4– Alert Resolution Effectiveness, as the quantity of accepted and canceled alerts stratified by risk rating.

Risk	Result	Traditional Alert	UCD Alert	p
D	Accepted	48	1	<0.01
	Canceled	61	79	
X	Accepted	38	18	0.21
	Canceled	23	21	

In Figure 3 we present the results of one of the questions from the satisfaction survey.

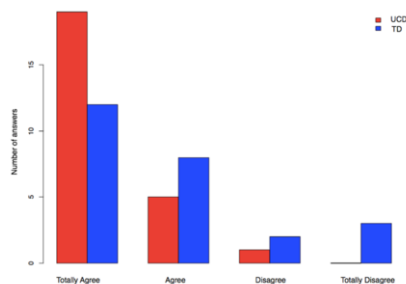


Figure 3—Satisfaction Survey Results regarding Alert Utility: “DDI alerts are useful for patient care” ($p = 0.03$). The bars represent the amount of answers for each degree of agreement. Red is for UCD and blue for traditional interface.

We performed 12 interviews, 5 were traditional and 7 UCD interface users. We analyzed their perceptions on the triggered DDI alerts. Utility perception differed regarding seniority, with better appraisal from junior physicians.

Discussion

This paper shows the fourth phase of a long-standing research project at the HIBA. The first phase implied the knowledge database redesign [19]. The second phase included two cycles of participatory design sessions, in which interface prototypes and evaluations focused on qualitative aspects [30]. Participants agreed that they wanted short, clear, and quick alerts [14,15]. A laboratory crossover study using clinical vignettes was the third phase. It showed quantitatively that our UCD method was a reliable way of designing and developing better DDI alerts. The results regarding effectiveness, efficiency, and user satisfaction were similar to those in Russ et al. [23]. Regarding efficiency, the new interface required less time to complete the task but the same amount of clicks and justification words as in the old one, probably due to quick and enhanced interaction with the alert. The UCD interface showed statistically significant improvements in effectiveness and user satisfaction. The interviews showed that nearly 60% of the users preferred the UCD interface.

Such promising results required a real scenario validation study. Regarding the alert override, our results were similar to previous publications that reported cancellation rates between 49% and 96% [9]. The global effectiveness was higher for the standard version, opposing our previous lab study results, in which the UCD interface performed better. We then performed a stratified analysis and detected that there were no significant differences in effectiveness regarding X risk rating (“avoid combination”). The D risk rating (“modify regimen”) had a slight trend for the standard version. The qualitative approach of surveys and interviews gave an insight on these results. The surveys were answered a long time after the alert exposure (weeks), thus they might reflect user attitudes towards the software instead of their opinions on the task itself. The new UCD alert scored better in the satisfaction survey than the traditional alert. From the interviews analyses, we discovered that senior staff relied more on their clinical experience, while junior physicians appreciated the benefits of the decision support. Both agreed about the utility of the new UCD alert, as it can prevent unintended errors especially in the context of urgencies, work overload, and time constraints. Resident physicians also used the CDSS as a learning opportunity. We

also realized that alerts usually fail to change the physician intention to prescribe. Thus, providers keep ordering the drugs despite potential interactions. Although, they monitored the drug effect and adverse events as they have been warned in advance. We named this medical awareness as “safe omission”. This might be underlying the high alert override rate found in many previous CDSS studies.

The research was done in a single academic center using in-house developed software and thus might not represent other institutions.

Conclusion

In real scenarios, user-centered design alerts were more usable, being more effective and satisfactory, but less efficient than traditional alerts. “Safe omission”, as a new concept, emerged from our stratified analyses and interviews.

It is necessary to continue the analysis of human-alert interaction, especially regarding the “safe omission” phenomenon. Further research in this field is recommended.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] L.T. Kohn, J.M. Corrigan, M.S. Donaldson, *To Err Is Human: Building a Safer Health System*, Institute of Medicine: National Academies Press, 1999.
- [2] L.L. Leape, D.W. Bates, D.J. Cullen, J. Cooper, H.J. Demonaco, T. Gallivan, R. Hallisey, J. Ives, N. Laird, G. Laffel, Systems analysis of adverse drug events. ADE Prevention Study Group, *JAMA*. **274** (1995) 35–43.
- [3] T.J. Bright, A. Wong, R. Dhurjati, E. Bristow, L. Bastian, R.R. Coeytaux, G. Samsa, V. Hasselblad, J.W. Williams, M.D. Musty, L. Wing, A.S. Kendrick, G.D. Sanders, D. Lobach, Effect of clinical decision-support systems: A systematic review, *Ann. Intern. Med.* **157** (2012) 29–43.
- [4] P. a Glassman, B. Simon, P. Belperio, A. Lanto, Improving recognition of drug interactions: benefits and barriers to using automated drug alerts., *Med. Care*. **40** (2002) 1161–71.
- [5] R. Greenes, *Clinical Decision Support: The Road Ahead*, Academic Press, 2014.
- [6] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S.C. Morton, P.G. Shekelle, Systematic review: impact of health information technology on quality, efficiency, and costs of medical care, *Ann. Intern. Med.* **144** (2006) 742–752.
- [7] J.J. Coleman, H. van der Sijs, W.E. Haefeli, S.P. Slight, S.E. McDowell, H.M. Seidling, B. Eiermann, J. Aarts, E. Ammenwerth, A. Slee, R.E. Ferner, A. Slee, On the alert: future priorities for alerts in clinical decision support for computerized physician order entry identified from a European workshop., *BMC Med. Inform. Decis. Mak.* **13** (2013) 111.
- [8] M. Xie, M.B. Weinger, W.M. Gregg, K.B. Johnson, Presenting Multiple Drug Alerts in an Ambulatory Electronic Prescribing System, *Appl. Clin. Inform.* **5** (2014) 334–348.
- [9] H. van der Sijs, J. Aarts, A. Vulto, M. Berg, Overriding of drug safety alerts in computerized physician order entry, *J. Am. Med. Informatics Assoc.* **13** (2006) 138–147.
- [10] D. Luna, F. Plazzotta, C. Otero, F. González Bernaldo de Quirós, A. Baum, S. Benítez, *Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires*, CEPAL, 2012.
- [11] D. Luna, M. Waldhorn, A. Ianicelli, M. Risk, F. González Bernaldo De Quirós, Drug-drug interaction in a healthcare network of Argentina: frequency and characteristics, in: XIII Congr. Bras. Em Informática Em Saúde – CBIS 2012, 2012.
- [12] Institute of Medicine (IOM), *Health IT and Patient Safety: Building Safer Systems for Better Care*, The National Academies Press (US), Washington (DC), 2012.
- [13] M.D. Paterno, S.M. Maviglia, P.N. Gorman, D.L. Seger, E. Yoshida, A.C. Seger, D.W. Bates, T.K. Gandhi, Tiering Drug-Drug Interaction Alerts by Severity Increases Compliance Rates, *J. Am. Med. Informatics Assoc.* **16** (2009) 40–46.
- [14] A. Feldstein, S.R. Simon, J. Schneider, M. Krall, D. Laferriere, D.H.

- Smith, D.F. Sittig, S.B. Soumerai, How to design computerized alerts to safe prescribing practices., *Jt. Comm. J. Qual. Saf.* **30** (2004) 602–13.
- [15] J.D. Duke, X. Li, P. Dexter, Adherence to drug–drug interaction alerts in high-risk patients: a trial of context-enhanced alerting, *J. Am. Med. Informatics Assoc.* **20** (2013) 494–498.
- [16] E. Ammenwerth, W.O. Hackl, D. Riedmann, M. Jung, Contextualization of automatic alerts during electronic prescription: Researchers' and users' opinions on useful context factors, *Stud. Health Technol. Inform.* **169** (2011) 920–924.
- [17] D.W. Bates, G.J. Kuperman, S. Wang, T. Gandhi, A. Kittler, L. Volk, C. Spurr, R. Khorasani, M. Tanasijevic, B. Middleton, Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-based Medicine a Reality, *J. Am. Med. Informatics Assoc.* **10** (2003) 523–530.
- [18] D.S. McEvoy, D.F. Sittig, T.-T. Hickman, S. Aaron, A. Ai, M. Amato, D.W. Bauer, G.M. Fraser, J. Harper, A. Kennemer, M.A. Krall, C.U. Lehmann, S. Malhotra, D.R. Murphy, B. O'Kelley, L. Samal, R. Schreiber, H. Singh, E.J. Thomas, C. V Vartian, J. Westmorland, A.B. McCoy, A. Wright, Variation in high-priority drug-drug interaction alerts across institutions and electronic health records, *J. Am. Med. Informatics Assoc.* **24** (2016).
- [19] D. Luna, V. Otero, D. Canosa, S. Montenegro, P. Otero, F. González Bernaldo de Quirós, Analysis and Redesign of a Knowledge Database for a Drug-drug Interactions Alert System, *Stud Health Technol Inform.* **129** (2007) 885–889.
- [20] R.T. Scheife, L.E. Hines, R.D. Boyce, S.P. Chung, J.D. Momper, C.D. Sommer, D.R. Abernethy, J.R. Horn, S.J. Sklar, S.K. Wong, G. Jones, M.L. Brown, A.J. Grizzle, S. Comes, T.L. Wilkins, C. Borst, M.A. Wittie, D.C. Malone, Consensus recommendations for systematic evaluation of drug-drug interaction evidence for clinical decision support., *Drug Saf.* **38** (2015) 197–206.
- [21] Lexicomp, Lexicomp, (2016).
- [22] H.M. Seidling, S. Phansalkar, D.L. Seger, M.D. Paterno, S. Shaykevich, W.E. Haeefeli, D.W. Bates, Factors influencing alert acceptance: a novel approach for predicting the success of clinical decision support, *J Am Med Inf. Assoc.* **18** (2011) 479–84.
- [23] A.L. Russ, S. Chen, B.L. Melton, E.G. Johnson, J.R. Spina, M. Weiner, A.J. Zillich, A Novel Design for Drug-Drug Interaction Alerts Improves Prescribing Efficiency., *Jt. Comm. J. Qual. Patient Saf.* **41** (2015) 396–405.
- [24] V.L. Patel, T.G. Kannampallil, Cognitive informatics in biomedicine and healthcare, *J. Biomed. Inform.* **53** (2015) 3–14.
- [25] D.A. Norman, S.W. Draper, *User Centered System Design: New Perspectives on Human-computer Interaction*, 1 edition, CRC Press, Hillsdale, N.J., 1986.
- [26] ISO. 92410-210:2010: *Part 210: Ergonomics of human–system interaction: Human-centred design for interactive systems.*, ISO, 2010.
- [27] A. Kushniruk, C. Nohr, Participatory Design, User Involvement and Health IT Evaluation., *Stud. Health Technol. Inform.* **222** (2016) 139–51.
- [28] A. Joshi, M. Arora, L. Dai, K. Price, L. Vizer, A. Sears, Usability of a Patient Education and Motivation Tool Using Heuristic Evaluation, *J. Med. Internet Res.* **11** (2009).
- [29] J. Simonsen, T. Robertson, *Routledge international handbook of participatory design*, Routledge, 2013.
- [30] D. Luna, C. Otero, A. Almerares, E. Stanziola, M. Risk, F. Gonzalez Bernaldo de Quirós, Participatory design for drug-drug interaction alerts, *Stud. Health Technol. Inform.* **210** (2015) 45–49.
- [31] S.C. Evans, M.C. Roberts, J.W. Keeley, J.B. Blossom, C.M. Amaro, A.M. García, C.O. Stough, K.S. Canter, R. Robles, G.M. Reed, Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies, *Int. J. Clin. Heal. Psychol.* **15** (2015) 160–170.
- [32] D. Luna, C. Otero, M. Risk, E. Stanziola, F. González Bernaldo de Quirós, Impact of Participatory Design for Drug-Drug Interaction Alerts. A Comparison Study Between Two Interfaces, **228** (2016) 68–72.
- [33] D.R. Luna, D.A. Rizzato Ledo, C.M. Otero, M.R. Risk, F. González Bernaldo de Quirós, User-centered design improves the usability of drug-drug interaction alerts: Experimental comparison of interfaces., *J. Biomed. Inform.* **66** (2017) 204–213.
- [34] E. Stanziola, M.Q. Uznayo, J.M. Ortiz, M. Simón, C. Otero, F. Campos, D. Luna, User-Centered Design of Health Care Software Development: Towards a Cultural Change, *Stud. Health Technol. Inform.* **216** (2015) 368–371.
- [35] K. Zheng, K. Fear, B.W. Chaffee, C.R. Zimmerman, E.M. Karls, J.D. Gatwood, J.G. Stevenson, M.D. Pearlman, Development and validation of a survey instrument for assessing prescribers' perception of computerized drug-drug interaction alerts, *J. Am. Med. Informatics Assoc.* **18** (2011) i51–i61.
- [36] SurveyMonkey, (n.d.).
- [37] R: The R Project for Statistical Computing, (n.d.).

Address for correspondence

Daniel R Luna, MD, MSc, PhD daniel.luna@hospitalitaliano.org.ar

Extracting Follow-Up Recommendations and Associated Anatomy from Radiology Reports

Thusitha Mabotuwana^a, PhD, Christopher S Hall^a, PhD, Sandeep Dalal^b, MS, Joel Tieder^c, MD, MPH, Martin L. Gunn^c, MBChB, FRANZCR

^aPhilips Healthcare, Seattle, WA, USA, ^bPhilips Research, Cambridge, MA, USA, ^cUniversity of Washington, Seattle, WA, USA

Abstract

Adherence rates for timely imaging follow-up are usually low due to low rates of diligence by referring physicians and/or patients with following recommendations for follow-up imaging. This can lead to delayed treatment, poor patient outcomes, unnecessary testing, and legal liability. Existing follow-up recommendation detection methods are often disease- and modality-specific. To address some of these limitations, we present a generic radiology report processing pipeline that can be used to extract follow-up imaging recommendations by anatomy using an ontology-based approach. Using a large dataset from three hospitals, we discuss our methodology in the context of identifying follow-up imaging recommendations that are related to lung, adrenal and/or thyroid conditions. The algorithm has 99% accuracy (95% CI: 95.8-99%). We also present an interactive dashboard that can be used to understand trends related to follow-up recommendations.

Keywords:

Medical Informatics Applications; Follow-Up Studies; Quality Assurance, Health Care

Introduction

Radiology reports often contain follow-up imaging recommendations for further diagnostic characterization or to ensure stability of a potentially malignant lesion [1]. Sometimes, these findings are not expected by the ordering provider and are commonly referred to as *incidental findings* since they are incidental to the reason for the current exam – for example, a small pulmonary nodule in the lower lobe of the lung may be detected on a computed tomography (CT) abdomen study that was ordered to rule out appendicitis. Failure to appropriately address imaging follow-up recommendations in a timely manner is common (in one study, 12% of cases of potential malignancy [2]) and can lead to delayed treatment, poor patient outcomes, unnecessary testing, lost revenue, and legal liability [1-4].

Imaging follow-up adherence is low and over 35% of follow-up imaging recommendations are usually not followed [5]. Clinicians may determine that follow-up is unnecessary, especially when a follow-up recommendation is conditional on clinical findings that were not available to the radiologist making the recommendation [1]. In one study, follow-up rates dropped from 78.8% for no conditional language to 43.8% when conditional language was present [6]. Various other

reasons have been attributed to failure to follow-up, including the referring physicians missing the recommendations or losing track while addressing a more acute illness, loss of information during handover between care teams, the recommendation not being communicated to the patient, and the patient failing to schedule or show-up for the follow-up appointment [7]. The first step towards ensuring timely follow-up of imaging recommendations is to detect presence of recommendations in radiology reports. Several researchers have explored how follow-up recommendations can be identified, but much of prior work has focused on identifying follow-up recommendations for a specific modality, such as CT, incidental findings only [8, 9] or a particular type of finding, such as pulmonary nodules [10] or adrenal masses [11]. In order for follow-up detection algorithms to be more useful in practice, they need to be scalable and generic so that recommendations can be identified from all radiology reports irrespective of modality and type of finding.

To address some of the limitations in currently available techniques, we present a radiology report processing pipeline that can be used to: (1) extract follow-up recommendations, and then (2) extract the anatomy related to the follow-up recommendation. Being able to associate a follow-up recommendation with anatomy has three main benefits. First, it allows for identification of different types of findings (e.g. lung nodules) to be recognized in a generic manner; second, it provides a mechanism to quantify follow-up recommendations by anatomy (e.g. all lung findings) which can be used to determine appropriate interventions to improve follow-up adherence, focusing on radiology sections where adherence is particularly poor; and third, the anatomy can be used as a surrogate to identify incidental findings by determining follow-up recommendations where the associated anatomy is not the same as the anatomy of ordered exam. The purpose of this study is to propose a novel pipeline to extract specific follow-up findings with their associated anatomic region.

<p>EXAMINATION: CT ABDOMEN AND PELVIS W CONTRAST</p> <p>CLINICAL INDICATION: Abdominal pain and distention.</p> <p>TECHNIQUE: CTB A1 CT Abdomen Pelvis Survey with IV contrast. Contiguous axial sections were obtained through the abdomen and pelvis with IV contrast only. Multiphase reformatted images were performed. Automated exposure control and statistical iterative reconstruction techniques substantially lowered patient radiation dose.</p> <p>CONTRAST: Omnipaque 350 mg/ml 121cc@3cc/sec ml mm/dd/yyyy 06:52 AM INTRAVENOUS</p> <p>COMPARISON: None.</p> <p>FINDINGS: Lung bases: Atelectasis/scarring in the middle lobe and lingula. Granuloma right lower lobe.</p> <p>ABDOMEN: Liver: Diffusely decreased density throughout the hepatic parenchyma, in keeping with fatty infiltration. Gallbladder and bile ducts: The common bile duct is dilated up to 1.5 cm. There are several high density foci within the distal common bile duct. There is mild intrahepatic biliary ductal dilatation. Pneumobilia is present. Spleen: Unremarkable. Pancreas: Pancreatic ductal dilatation, involving the proximal body and head, measuring up to 0.7 cm. Adrenals: 1.3 cm left adrenal nodule. Kidneys and Ureters: Hypodensities right kidney, the larger of which is a cyst; the smaller is too small to characterize. Cortical irregularity involving both kidneys, likely related to remote scarring.</p> <p>Abdominal Lymph Nodes and Retroperitoneum: No enlarged lymph nodes. Vasculature: Atherosclerotic calcification of the abdominal aorta and its branch vessels. Bowel: Scattered colonic diverticula without CT evidence of diverticulitis. Normal appendix right lower quadrant. There is a 1.4 cm hypodense jejunal lesion (3/101). Peritoneum and Abdominal Wall: Unremarkable.</p> <p>PELVIS: Pelvic Viscera: Apparent thickening of the endometrium. Bladder: Unremarkable.</p> <p>BONES: Multilevel degenerative changes. Mildly sclerotic focus L3 may represent a bone island.</p> <p>IMPRESSION: 1. Intra and extra hepatic biliary ductal dilatation with associated pancreatic ductal dilatation. High density within the distal common bile duct may represent choledochal stones with associated obstruction resulting in described ductal dilatation. Pneumobilia is presumably related to prior ERCP. Recommend GI consultation. <u>Consider ERCP for further evaluation.</u> 2. Indeterminate 1.3 cm left adrenal nodule. In the absence of a history of malignancy, this is likely a benign finding. <u>This could be assessed further if MRI is performed.</u> 3. Hypodense 1.4 cm jejunal lesion is of indeterminate etiology. <u>As this cannot easily be visualized endoscopically, follow-up outpatient small bowel follow-through and/or CT abdomen/pelvis in 1-6 months recommended.</u> 4. Apparent endometrial thickening is atypical for patient age. <u>Recommend outpatient pelvic ultrasound for evaluation.</u> 5. Hepatic steatosis.</p> <p>Urgent findings discussed with Dr. xx by Dr. xx on mm/dd/yyyy at 7:55 AM.</p>

Figure 1 – Sample radiology report with multiple follow-up imaging recommendations

Methods

Dataset and Strategy

We extracted 417,448 radiology reports generated between 1 January 2015 and 31 May 2016 from the University of Washington radiology information system for three network hospitals. For each report, several meta-data fields were also extracted, including exam date, radiology subspecialty, patient class and modality. The Human Subjects Division at the University of Washington determined that the study was IRB exempt as part of a quality improvement project.

The first step in the process was to identify reports that contained a follow-up recommendation. This was performed using a previously developed follow-up detection algorithm which parses the radiology report to extract sections (e.g., “Clinical Indication”, “Findings” and “Impression” as shown in Figure 1), paragraph headers within each section if any (e.g. “Abdomen” and “Pelvis”) and the sentences within the paragraphs. The algorithm then evaluates the sentences within the “Findings” and “Impression” sections to determine if a sentence contains a follow-up recommendation (e.g. “Given history of malignancy, follow-up CT chest in 3 months is recommended”). Follow-up detection is performed using keyword searches and other heuristics. The output of this first step is a list of follow-up recommendation sentences as shown underlined in Figure 1 (along with meta-data, such as whether

it is a negated sentence – e.g. “no further follow-up is necessary”). Using 532 reports annotated by three radiologists (including author MG) as the ground truth, the detection algorithm was evaluated to have 93.2% PPV (95% CI: 89.8-94.5%), 99.5 NPV (95% CI: 98.4-99.9%) and 97.9% accuracy (95% CI: 96.2-98.5%).

Next, to identify anatomy associated with the follow-up recommendation, we used an ontology based natural language processing engine we developed internally [12] along with publicly available NCBO annotation service [13]. Queries to both services were constrained to extract anatomies as defined by SNOMED-CT ontology. Results were then merged and unique values selected. This approach was selected to optimize the capabilities of the two systems, for instance, if the text contains “right lower lobe”, the anatomy engine would detect “Structure of right lower lobe of lung” corresponding to SNOMED ID 266005 whereas NCBO would not find a mapping. Conversely, from the sentence “hypervascular liver lesion, MRI follow-up is suggested”, NCBO detected “Liver Structure”, corresponding to SNOMED ID 10200004 whereas the internal engine identified “Lesion of liver” (SNOMED ID 300331000), which is a finding. Since our focus is on identifying anatomy, in this instance, the engine did not find any relevant anatomy since a longer phrase was matched.

Our follow-up anatomy detection algorithm was developed such that it first attempts to extract anatomy from the follow-up sentence – for instance, concept “Thoracic Structure” corresponding to ID 51185008 will be extracted from “Follow-up CT chest is recommended”. If no anatomy is identified in

Table 1 – Extracted anatomy for several follow-up recommendation sentences. Detected follow-up sentence is italicized

Anatomy Context	Extracted Anatomy	SNOMED-CT Description(s) and ID(s)
<i>These can be reassessed on CT <u>chest</u> for lung nodule follow-up.</i>	chest lung	Thoracic Structure, 51185008 Entire lung, 181216001
<i>There is a <u>right adrenal nodule</u> which is likely benign and could be further evaluated by CT at the time of lung nodule follow-up.</i>	right adrenal lung	Entire right adrenal gland, 281625001 Entire lung, 181216001
<i>1 cm hypoechoic focal lesion in the mid portion of the <u>left kidney</u>. Although it is possible that it may represent a simple cyst, it is not adequately characterized on this study. Recommend follow up US in 6 months to establish stability.</i>	left kidney	Left kidney structure, 18639004
<i>Nodular opacities in the <u>right lung</u> may represent infection versus aspiration. Dedicated CT may be helpful.</i>	right lung	Right lung structure, 3341006

the follow-up sentence, the algorithm steps backwards from the follow-up sentence, processing one sentence at a time, until at least one anatomy is identified in a sentence. Search was restricted to the section follow-up sentence occurred in (which is usually ‘Findings’ and/or ‘Impression’ sections). Once identified, the ‘anatomy context’ becomes the text from the beginning of matched sentence to end of follow-up sentence. This process was repeated for all follow-up sentences when a report contained multiple recommendations. Table 1 shows four examples of extracted anatomy. For each follow-up recommendation, we also keep track of the previous two sentences which is referred to as ‘search context’. This *search context* can then be queried using regular expressions to detect the type of follow-up (e.g. whether follow-up recommendation is for a pulmonary nodule).

A quality improvement oversight committee composed of multiple clinical and quality stakeholders focused first on three commonly occurring findings for which published follow-up guidelines exist: lung, thyroid, and adrenal nodules. Consequently, the scope of current research identified follow-up recommendations for these three findings. Keywords ‘nodule’, ‘lesion’, ‘tumor’, ‘lump’, and ‘mass’ were included for all three, while several additional descriptors were included at a finding-specific level. We included ‘opacity’ for lung findings; ‘hypodensity’ and ‘fullness’ for adrenal findings; and ‘hypodensity’ and ‘opacity’ for thyroid findings. We required one of these nodule-related words to be within a 6-word proximity (after removing stop words) of where the anatomy was detected within the anatomy context to ensure the finding was truly related to the detected anatomy.

Table 2 – Exams by type of follow-up recommendation

Type of Follow-up Recommendation	Number of Exams (n = 5547)	Percent of Exams with Follow-up Recommendation
Lung	4154	74.89%
Thyroid	641	11.55%
Lung and Thyroid	364	6.56%
Adrenal	320	5.77%
Lung and Adrenal	61	1.10%
Thyroid and Adrenal	6	0.11%
Lung, Thyroid, and Adrenal	1	0.02%

Table 3 – Exams with follow-up findings by scanned modality (‘Others’ include CR/DR, NM, US and interventional radiology)

Modality Associated with Follow-up	Number of Exams (n = 5547)	Percent of Exams with Follow-up Recommendation
CT	4265	76.89%
CR	590	10.64%
US	302	5.44%
MR	188	3.39%
Others	202	3.64%

To validate our algorithm, we manually selected 50 reports for each of the three follow-up finding types as well as 50 reports that contained a follow-up recommendation but were unrelated to the lung, adrenal or thyroid. This was performed by searching for the specific finding types in the “Findings” and “Impression” sections of randomly selected reports and repeating the process until the required dataset of 200 reports was created. The algorithm performance was 98.7% sensitivity (95% CI: 96.5-98.7%), 100% specificity (95% CI: 93.6-100%), and 99% accuracy (95% CI: 95.8-99%). There were two false-negatives, one related to an adrenal nodule and the other related to a lung nodule. A false-negative was defined as an instance where follow-up detection or anatomy extraction failed. Overall accuracy was slightly better than 97.9% follow-up detection accuracy reported previously since detection errors were rectified prior to anatomy extraction, which is the focus of the work presented herein.

Results

There were 67,521 (16.2%) reports that had at least one follow-up imaging recommendation sentence. Of these, 5,547 exams contained at least one of the three specific finding types of interest. Table 2 shows distribution of the follow-up recommendations by finding type and Table 3 shows distribution by scanned modality. There were 4,265 CT Chest exams for all anatomies that contained a follow-up recommendation of which 3,451 (80.9%) were lung related.

In order to provide radiology administrators with a quick overview of departmental trends and ability to monitor effectiveness of QI interventions over time, we also developed an interactive dashboard (using Microsoft Power BI [14], Microsoft Corporation, Redmond WA) that shows the number and rate of reports containing follow-up sentences by anatomy, modality as well as by month. Various filters have

been provided so that the data can be explored to understand trends and identify barriers to improvement. Figure 2 shows number of follow-up reports containing recommendation sentences that are lung, thyroid or adrenal related for all sections across all three hospitals for the entire duration. A user can easily explore the underlying data that contributes towards a particular metric, for instance, a user can right-click on the CT bar showing 4.26k and examine specific reports.

Discussion

In this paper we have outlined a generic pipeline that can be used to extract follow-up recommendations and their associated anatomy using an ontology-based approach. Using production data, we have demonstrated how the pipeline can be used to extract follow-up recommendation sentences associated with lung, thyroid, and adrenal nodules for multiple imaging modalities. Using the follow-up anatomy detection capabilities, radiology administrators can determine which radiology sections make more follow-up recommendations (as a percentage) and design appropriate interventions to improve follow-up adherence rates. Further, the technology has the potential to be used as a surrogate to identify incidental findings by filtering for exams where the anatomy of follow-up recommendation is different from anatomy of ordered exam. The generic nature of the anatomy and finding extraction algorithm could also be more scalable compared to some of the prior work where the primary focus was on identifying follow-up recommendations for a specific modality, patient setting or a particular type of finding [3, 8-11, 15]).

Despite using a large dataset from three institutions, the current study has limitations. First, all reports were created using common dictation macros (with somewhat limited expressiveness compared to fully free text reports) that are shared across



Figure 2 – Dashboard showing exams with follow-up recommendations

the network hospitals and therefore the methods used to parse radiology reports may not be readily generalizable to other institutions. Second, the algorithm performed imperfectly in 2 out of the 200 reports we examined. In one of these instances, the algorithm missed the follow-up statement which was mentioned in conjunction with another (“The attenuation coefficient of the left adrenal nodule is about 10 Hounsfield units. Therefore, it cannot be characterized as an adenoma. This could be characterized by CT at the same time as a renal mass protocol”). Although the pipeline failed to recognize this recommendation, it did identify the follow-up recommendation for the renal mass in the previous sentence, potentially ensuring that follow-up would occur. In the other failed instance, “Multiple gray nodules are unchanged in size compared to prior, but remain indeterminate. Recommend follow-up CT in 12 months to assess for stability”, the follow-up recommendation was correctly detected, but the anatomy was not (‘gray nodules’ does not match any anatomy concepts in SNOMED). In fact, this also shows some of the limitations of using an ontology-based approach to detect anatomy. Complementing the ontology-based approach with domain-specific dictionaries (e.g. using a text-to-anatomy dictionary where “gray nodule” is a key used to refer to anatomy “lung”) could be one option. We are also looking into generalizing the detection of nodule-related concepts. For instance, filtering for ‘morphologic abnormality’ concepts in SNOMED could be an option instead of specifying variants for ‘nodule’.

There are several potential applications for this algorithm in our department. First, it can be used as a means to improve precision of follow-up recommendation sentences. We have found that a significant proportion of follow-up sentences did not contain the follow-up imaging modality (18%), a recommended timeframe (67%), or both (11%). These incomplete follow-up recommendation sentences can be confusing for ordering clinicians who are unfamiliar with published imaging follow-up guidelines. We are in the process of implementing standardized follow-up macros and ensuring guideline uniformity among radiology sections in our department, and plan to measure the impact using this pipeline. Second, we can use this algorithm to benchmark our own compliance rate and variability with published clinical and departmental follow-up guidelines. Third, we can use this algorithm to ensure timely follow-up imaging of the appropriate body region using the appropriate modality.

To improve follow-up compliance and appropriateness, it is important to identify follow-up recommendations with their associated anatomic location. We have demonstrated that a robust pipeline can be developed to identify lesions of the lung, thyroid, and adrenal glands to measure quality of the follow up recommendation and to ensure timely follow-up. Given the large number of anatomic sites and diseases that are associated with these recommendations, it is important to have techniques that are easily extensible without excessive human rework. The proposed generic report processing pipeline shows potential to be used as one such approach.

Acknowledgements

The authors would like to acknowledge the contributions of Erik Christiansen (Radiology Administrator) for all his support in providing access to radiology reports from RIS.

References

- [1] H.B. Harvey, C.C. Wu, M.D. Gilman, et al., Correlation of the strength of recommendations for additional

- imaging to adherence rate and diagnostic yield, *J Am Coll Radiol* **12**(10) (2015), 1016-1022.
- [2] C.E. Sloan, S.C. Chadalavada, T.S. Cook, C.P. Langlotz, M.D. Schnall, and H.M. Zafar, Assessment of follow-up completeness and notification preferences for imaging findings of possible cancer: what happens after radiologists submit their reports? *Acad Radiol* **21**(12) (2014), 1579-1586.
- [3] S. Dutta, W.J. Long, D.F. Brown, and A.T. Reisner, Automated detection using natural language processing of radiologists recommendations for additional imaging of incidental findings, *Ann Emerg Med* **62**(2) (2013), 162-169.
- [4] R. Kreisman, \$2.5 Million settlement for doctor's failure to follow up on signs and symptoms of lung cancer. [cited 2016 Nov 8]; Available from: <http://www.robertkreisman.com/medical-malpractice-lawyer/2014/02/2-5-million-settlement-doctors-failure-follow-signs-symptoms-lung-cancer.html>.
- [5] J.L. Callen, J.I. Westbrook, A. Georgiou, and J. Li, Failure to follow-up test results for ambulatory patients: a systematic review, *J Gen Intern Med* **27**(10) (2012), 1334-1348.
- [6] M.L. Gunn, B.E. Lehnert, C.S. Hall, M. Yetisgen, K. Trovato, and S. Dalal, Use of conditional statements in radiology follow-recommendation sentences: relationship to follow up compliance. Radiological Society of North America 101st Scientific Assembly and Annual Meeting, Chicago, 2015.
- [7] M. Kulon, Lost to follow-up: automated detection of patients who missed follow-ups which were recommended on radiology reports. Society for Imaging Informatics in Medicine Annual Meeting, 2016.
- [8] P.M. Jairam, M.J. Gondrie, D.E. Grobbee, et al., Incidental imaging findings from routine chest CT used to identify subjects at high risk of future cardiovascular events, *Radiology* **272**(3) (2014), 700-708.
- [9] D. Verdini, A.M. Lee, A.M. Prabhakar, S. Abbara, and B. Ghoshhajra, Detection of cardiac incidental findings on routine chest CT: the impact of dedicated training in cardiac imaging, *J Am Coll Radiol* (2016), 1-5.
- [10] D.P. Blagev, J.F. Lloyd, K. Conner, et al., Follow-up of incidental pulmonary nodules and the radiology report, *J Am Coll Radiol* **11**(4) (2014), 378-383.
- [11] J.J. Zopf, J.M. Langer, W.W. Boonn, W. Kim, and H.M. Zafar, Development of automated detection of radiology reports citing adrenal findings, *J Digit Imaging* **25**(1) (2012), 43-49.
- [12] S.A. Hasan, X. Zhu, Y. Dong, J. Liu, and O. Farri, A Hybrid Approach to Clinical Question Answering. Twenty-Third Text Retrieval Conference, Maryland, 2014.
- [13] National Center for Biomedical Computing, NCBO Annotator. [cited 2016 September 5]; Available from: <https://biportal.bioontology.org/annotator>.
- [14] Microsoft, Power BI. [cited 2016 Nov 10]; Available from: <https://powerbi.microsoft.com>.
- [15] M. Yetisgen-Yildiz, M.L. Gunn, F. Xia, and T.H. Payne, Automatic identification of critical follow-up recommendation sentences in radiology reports, AMIA Annu Symp Proc, (2011), 1593-1602.

Address for correspondence

Thusitha Mabotuwana
Philips Healthcare, 22100 Bothell Everett Hwy, Bothell, WA 98021,
USA. thusitha.mabotuwana@philips.com

BarCode Medication Administration in ICU: Learning from Our Nurses

Thelma Maydana, Liliana Giraldo, Zulma González,
Bibiana Schachner, John Mayan, Daniel Luna, Sonia Benítez

Department of Health Informatics, Hospital Italiano de Buenos Aires

Abstract

Errors during drug administration stage represent a significant percentage of adverse events associated with health care. Its prevalence is higher in critical care units, due to vulnerable patient population, characteristics of intensive care as well as complexity of pharmacotherapy. Errors can occur at any stage of the medication cycle although they often occur during administration, so nurses' role in prevention is essential. The bar code medication administration (BCMA) technology increases patient safety by instituting control time, but a poorly designed system or a faulty implementation can lead to low utilization in future. This paper describes implementation, evaluation, detection of problems and search for their possible solutions of a BCMA system in the intensive care unit at Hospital Italiano de Buenos Aires.

Keywords:

Patient Safety; Medication Systems; Medication Errors

Introduction

Patient safety has become a priority for health care systems around the world. Several studies have shown that morbidity and mortality caused by drugs are very high. Even more alarming issue is that this problem is largely due to errors or faults that occur during hospitalization [1-2]. (Leape, 1995) and (Bates et al., 1998) also indicated that drug-preventable errors occur more frequently in prescribing (56%) and administration (34%), and to a lesser extent in transcription (6%) and dispensation (4%). The easily intercepted errors were those that occurred in the early stages of the process, specifically prescription (48%) – of which 29% were due to lack of knowledge of the drug or that of the patient [3-4]. Other authors indicate that errors occur more frequently in the stages of preparation and administration of drugs, as well as 78% of medication errors leading to an adverse effect are due to failures in the prescription-dispensation-administration circuit, which could be optimized with the use of computerized information systems [7-8]. In intensive care unit (ICU), medication errors can occur in one-third of hospitalized patients and have the potential to cause permanent damage to patients, prolong hospital stay, and add to the associated emotional and financial costs [3-4]. Preparation and administration of drugs are the sole responsibility of nursing staff in most health organizations and hence we need to emphasize the importance of the safety culture. Furthermore, there are multiple causes for errors in preparation and pharmacological administration in hospital environment, such as: overload, lack of knowledge, human errors including inattention and defective work processes [9].

The bar code medication administration (BCMA) technology represents an additional strategy to prevent drug errors in the administration stage and an improvement in patient safety

[3][6]. These systems have been shown to reduce errors of drug administration when used as a closed loop system, where drug prescription, dispensing, and drug delivery processes are electronically linked [4] [6].

The BCMA system makes the drug administration process safer by reading the barcode on patient's identification bracelet and that of the medication pouches to ensure that the correct medication, in the correct dose, is administered to the right patient in the right way and at the right time [10]. The processes intermediated with barcode control, such as medication administration, contribute to patient safety through barcode technology and real-time network connectivity and are used to improve the accuracy of drug administration. This system is currently limited to environments equipped with computers, barcode readers at bedside, and compatible information systems [11-12].

The objective of this work is to describe implementation of a BCMA system and evaluation of its use in the ICU at Hospital Italiano de Buenos Aires (HIBA).

Methods

Setting

Hospital Italiano de Buenos Aires (HIBA), in Argentina, is an institution with 150 years of history, with over 2700 physicians, more than 1200 nurses, and 1800 administrative and support employees. It is a university hospital that covers the entire spectrum of health care from outpatient care, emergencies, acute care, medical and surgical specialties, critical care, home care, and chronic care. It encompasses a network of two hospitals with 750 beds (200 for intensive care), 41 operating rooms, 800 home care beds, 25 outpatient clinics, and 150 associated private practices located in Buenos Aires city and its suburban area.

The institution has been designing and building its own health information system since 1998 that includes clinical and administrative data. Its Electronic Health Record (EHR) system is an integrated, modular, problem oriented, and patient centered system that works in different clinical settings and allows computer physician order entry for medications and medical tests, storage and retrieval of tests results and images, and communication system. It is Joint Commission International (JCI) accredited and has been recently certified by the HIMSS as level 6+ in the Electronic Medical Record Adoption Model (EMRAM), being the first hospital in Argentina and the second in Latin America reaching this stage. The electronic nursing record was developed in stages since 2010. It is integrated in the EHR and organized by sections: Assessment, Diagnosis, Planning, Implementation and Evaluation. Nurses are expected to follow the nursing process logic to document the care provided.

The intensive care unit (ICU) consists of 38 beds, divided into four sectors according to a severity criteria and therapeutic requirements of the patients. Rooms are individual for each patient and each one has a computer on wheels dedicated to the room. 100 nurses work in the area, distributed in 5 shifts (morning, afternoon, night A, night B, and weekend).

Design

Descriptive, observational study with quantitative and qualitative data analysis.

Area selection

Situation diagnosis: Different inpatient scenarios were surveyed during 2015, taking into account nurse-patient ratio, workstation locations, displacement spaces, and feasibility of diverse devices (computer/mobile). Finally, the ICU was selected because it had a specific bunker for preparation of medications, and also the ICU rooms are individual, each having a computer.

Equipment: Each room was equipped with a barcode reader with a usb cord because the distance between the computer and the patient was short (3-5 meters) and the usb type was less expensive than the wireless one. The area (called a bunker) for preparation of medications was equipped with label printers to identify medications.

Software Development

New Features: A medication preparation work list was initially developed in the EHR, where the medication preparer can filter the medication according to its sector, patient, route and schedule and print medication identification labels with QR codes ("Quick response" codes, a type of two-dimensional barcode) which contain the ID of the medical prescription, patient ID, drug, dose, route, and schedule. At the time of the medication administration, the data are checked against the identity of the patient by reading an identifying bracelet containing the data in the QR code. In addition, a new functionality was developed to record administration through barcoding, only visible in the EHR in the Adult Intensive Care Unit (ICU) and embedded in the administration section of the nursing e-chart. The barcode scanning of drugs is not available for 100% of the drugs administered to patients and hence the possibility of manually recording them remained available. This way nurses can choose the method they want to use.

Implementation

In-service Education and Training: Informatics nurses and physicians trained ICU nurses for all shifts. The bar-code scanner use and the new EHR feature were demonstrated. The new functionality was launched in April 2015.

Support: Support was provided in the sector during the first month after implementation to answer questions arising from the use, solve problems associated with the scanners and codes, and to train nurses who had been absent during the in-service training. In addition, the nurses had an institutional help desk for ordering.

Evaluation

Usage Rates: Queries were made to the database to know the barcoding (BC) frequencies in comparison to the manual records in the EHR in the following months after implementation.

The use of BC was always below the expected values compared to the manual ones, taking into account the drugs available for barcode scanning in our institution.

Survey and Operating Groups: An evaluation was made using the survey technique and operating groups (OG) for data

collection in November of 2015. Two meetings were conducted based on a convenience sample with 30 ICU nurses. They were given a semi-structured questionnaire to evaluate how they felt about the BCMA system that asked about: ease of use, weight of scanner, system utility, interference with patient care, training, support, level of satisfaction with the system, and suggestions. The coordinators of the evaluation then proposed group reflection and discussion activities based on topics related to: a) BCMA benefits, b) problems associated with its implementation in the ICU, c) knowledge of process steps, and d) technical support.

The activities were carried out during working hours and in both meetings the nurses were separated into subgroups of 4-5 people who were given a paper and pen to take notes.

The data analysis included descriptive statistics and content analysis for observations and suggestions.

Results

248,091 records of medication administration were made in the EHR during 2015 – of which 63,741 were made by BC and 184,350 were made manually. 249,238 records of administration were made during 2016 – of which 118,059 were made by BC and 131,179 were made manually. Figure 1 shows the number of administrations with both methods by month for both the years. There are more records of administration carried out manually during 2015 versus with barcode scanners, with the exception of September 2015 when the records using BC were slightly higher. During 2016, there was an increase in BC records in the month of April and the period from July to October, with a decrease in November.

21 nurses answered the questionnaire. 58% had been working in the hospital for 1-10 years and 33% between 11-20 years. 62% belonged to the shift 'morning', and the rest from afternoon shift. 50% were within the range of 30-39 years of age. In addition, 76.2% had a bachelor's degree in nursing. 18 participants (95%) indicated that the BCMA system was Very easy/Somewhat easy to understand and 15 (79%) reported that it was Very easy/Somewhat easy to use. Regarding the ease of scanning the bar-codes, 13 respondents (68.5%) indicated that it was Very easy/Somewhat easy and 2 persons noted that it was Somewhat difficult (10.5%). 17 of the respondents (90%) Agreed/Somewhat agreed that the BCMA system was useful for their work. Regarding interference with patient care (awakening or discomfort), 7 nurses (37%) considered that it interfered with the patient, while 53% (10) indicated that it did so Little/Nothing. At the same time, 18 (95%) participants indicated that the system reduces the probability of medication errors. In addition, 14 nurses (77.8%) indicated that they had received training to use the system while 4 responded that they did not. 56.3% (11) indicated that system failures received technical support to solve it, while 37% (7) said no. Of the total respondents, 58% indicated that they were satisfied with the use of BCMA for medication administration while the remaining (42%) were moderately satisfied.

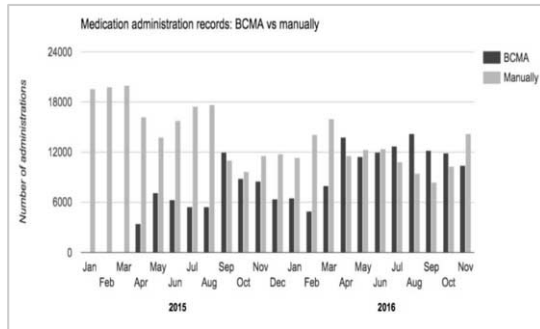


Figure 1– BCMA vs manually administrations

Three themes were identified related to: hardware, software, and process and infrastructure (including sub themes like training and technical support). Problems related to bracelets and difficulties in scanning the codes were reported. Other obstacles identified were: the slowness of EHR when the nurses choose to do administration with BC, the wires of usb scanners that drag over the floor and then pass over the patient's bed, which makes them unhygienic or they are unplugged when scanning the identification bracelets, so the process must be started all over again. Also the impossibility of scanning some medication such as the so-called 'multi dose' emerged, and those related to changes of the daily programmed medical indications. The most important findings are presented in tables 1-3:

Table 1– Hardware

Problems
'We must scan several times until we read the code (three attempts)'
'The legend "wrong patient" appears and does not decrease with the wristband replacement'
'The patient, the bracelet and the medications are correct and still shows that 'is not correct' and you have to reset the computer'
'Medications code is printed cut out'

Table 2– Software

Problems
'The system fails'
'The system slows down'
'After 5 PM, once the medical prescriptions' schedule changed, if you want to scan a drug label printed just before, you cannot, says incorrect'
'The bracelets are changed systematically on Tuesdays and Fridays but reading does not improve'
'The BCMA system does not record the minutes'

Discussion

The objective of this work was to describe the implementation of barcode technology for medication administration as well as the evaluation of its use in the adult ICU. 18 months after implementation, BCMA records continue to be lower than those manually recorded. They could not be 100% since we do not have the possibility to scan all medication labels, a factor to consider when implementing a BCMA system [13]. In those months where improvements were appreciated, it is due to greater monitoring and support in the area, associated with the

processes of accreditation we were going through, similar to those during JCI audit (September, 2015) and the EMRAM-HIMSS certification process (April, 2016). Under such circumstances, each problem was solved by assigning specific personnel to the task. Also problematic computer equipment was replaced immediately, a practice recommended to avoid decreasing nurses' productivity while waiting for the equipment to be repaired [14]. But when that support is removed and supervision goes down, so does the barcode use. The reasons reported by nurses who give up its use are related to the findings of the evaluation, such as: inadequate reading of codes in the face of repeated attempts, failure of the scanners or the computer, delayed response of the help desk, discomfort caused to the patients, drugs that cannot be scanned. A mixed process (barcode and manual) slows the workflow for nurses, who then choose to do it only manually or by using the so-called 'shortcuts' [15], such as using a larger size code printed on a paper, instead of scanning it directly from the patient bracelet. Lack of improvement in the frequency of use of the BCMA system is probably because the reported problems have not yet been solved.

Table 3– Process and infrastructure

Findings and problems

In-service Training

'Reached with the training to use it'

Patient's Unit

'The computer carts are moved a lot and the wire is disconnected'

The wire:

- 'It is uncomfortable'
- 'Is very long or is very short (without extender)'
- 'You may stumble (insecure)'
- 'Gets dirty and then goes over the patient's sheets'

'When the ICU was built it was not known that we would end up using BCMA, and the distance from the computers to the patient was not contemplated when installing them. Hence the idea of having wireless scanners'

Patient's discomfort

'When the patient is lucid and awake it is uncomfortable, it is annoying for the patient'

'It awakens him'

'If the patient sleeps, the second time it fails I do it manually'

Support

'They come sometimes'

'We ask for a 'help desk' but we do not have time'

'The help desk response is not immediate'

'I do it manually and then I ask for help desk'

'The help desk calls you after two days, is not immediate'

Miscellaneous

'During urgencies' we cannot record the medication'

'Some medication like syrups, creams, we cannot scan them (they have no code)'

'It's easy to use'

We do not have a support system to date that responds immediately to a failure to read a code, computer failure, or EHR freezes – this is a recommendation reported by the literature as good practice [14]. Our BCMA system has not

achieved the expected adoption, however, the attitude of the nurses in the ICU regarding the system as a whole is positive and they consider it useful and easy to use; an aspect that shows positive attitude towards the acceptance of a new technology [16].

We learned that if we want our nurses to adopt the BCMA system in their practice, we must improve it, looking for a solution to each previously reported problem [17].

We are currently reviewing the bracelet printing circuit implemented eight years ago for this purpose. We are also working to adapt the drugs labels that cannot yet be scanned by our BCMA system, such as "multi doses" (ointments, aerosols, syrups) and some "single dose" preparations (insulin) and evaluating the possibility of replacing scanners with USB cord with wireless ones. In addition, as the HIBA is working on the development of barcode reading for other products in addition to medication, such as human milk, blood, enteral and parenteral nutrition, a specific help desk for barcoding based systems is also being considered.

Conclusion

The adoption of a bar code reading system by nurses to verify patient identity and administer medication is influenced by different factors. The evaluation of all parts of the process, identification of associated problems and working on solutions according to recommendations could create greater adherence to barcode use and could have positive impact on patient safety.

Acknowledgements

The authors would like to thank ICU nurses and supervisor, Nursing Department and Health Informatics Residency Program for the collaboration.

References

- [1] Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med* 1991; 324:370-6.
- [2] Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000; 38:261-71.
- [3] Leape LL, Bates DW, Cullen DJ y cols: Systems analysis of adverse drug events. *JAMA* 1995; 274: 35-43.
- [4] Bates DW, Leape LL, Cullen DJ, laird N, Petersen LA, Teich JM y cols: Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998; 280: 1311-16
- [5] Silva AEBC, Reis AMM, Miaso AI, Santos JO, Cassiani SHB. Adverse drug events in a sentinel hospital in the state of Goiás, Brazil. *Rev Latino-Am Enfermagem*. 2011 Mar - Apr; 19 (2): 378 -86.
- [6] Franco JN, Ribeiro G, D'Innocenzo M, Brícia PA. Perception of the nursing team about causes of errors in the administration of medication. *Rev Bras Enferm*. 2010 Nov - Dec; 63(6): 927 - 32.
- [7] Raschke RA, Gollhare B, Wunderlich TA, Guidry JR, Leibow AI, Peirce JC, y cols: A computer alert system to prevent injury from adverse drug events. Development and evaluation in a community teaching hospital. *JAMA* 1998; 280: 1317-1320. [Links]
- [8] Bates DW, Teich JM, Lee J, Seger D, Kuperman GJ, Luf NM, Boyle D, Leape L: The impact of computerized physician order entry on medication error prevention. *Journal of American Medical Informatics Association* 1999; 6:313-321.
- [9] Garjón J, Gorricho J. Seguridad del paciente : cuidado con los errores de medicación. *Serv Navarro Salud*. 2010;18(3):32-46.
- [10] van Onzenoort HA, van de Plas A, Kessels AG, Veldhorst-Janssen NM, van der Kuy PH, Neef C. Factors influencing bar-code verification by nurses during medication administration in a Dutch hospital. *Am J Health Syst Pharm*. 2008 Apr 1;65(7):644-8.
- [11] Gann M. Bar Code Technology in Healthcare. *Nursing (Lond)*. 2015;45(3):60-6.
- [12] Brockman J, Skinner A. Errors Prevented by and Associated with Bar-Code Medication. *Jt Comm J Qual Saf*. 2007;33(5):293-301.
- [13] Bar-coded medication labeling: setting the stage for bar-code-enabled point-of-care systems.
- [14] Patterson ES, Rogers ML, Render ML. Fifteen best practice recommendations for bar-code medication administration in the Veterans Health Administration. *Jt Comm J Qual Saf*. 2004 Jul;30(7):355-65
- [15] Puckett F: Medication-management component of a point-of-care information system. *Am J Health-Syst Pharm* 52:1305-1309, Jun. 15, 1995
- [16] Davis, FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 3 (1989) 319-340.
- [17] Talierecio V, Schachner B, Borbolla D, Luna D, Villalba E, Quiros F. The expectations of nurses about the implementation of a Barcoded Medication Administration System: a qualitative study. *Stud Health Technol Inform*. 2014;205:191-5.

Address for correspondence

thelma.maydana@hospitalitaliano.org.ar

liliana.giraldo@hospitalitaliano.org.ar

Procedure for Reconstruction of a Predictive Score of Severe Deterioration in Inpatients

Navas Hernán and Bourdin Elián

Department of Health Informatics, Sanatorio Finochietto, Buenos Aires, Argentina

Abstract

Adequate recording of patients' physiological parameters provides vital information for health teams. Up to 80% of general admission patients have physiological parameters outside the normal range within 24 hours prior to transfer to the intensive care unit. Early warning systems analyze these data with the aim of detecting early deterioration in patients, which support medical decision making for reducing morbimortality. There are more than 100 systems for detecting patients at risk. Particularly NEWS has a greater ability to discriminate patients at risk of cardiac arrest, unplanned admission to ICU, and death. Sanatorio Finochietto has a commercial Health Information System. In order to exploit the available information, an algorithm capable of constituting a score value associated with NEWS has been developed from the combination of measurements made at different times. The reconstruction procedure allowed a greater number of full records. We had a 75% coincidence with real-time comparison.

Keywords:

Electronic Health Records; Clinical Decision-Making; Computing Methodologies.

Introduction

Inpatients are always at risk of suffering clinical deterioration, usually characterized by the alterations in their physiological state. Up to 80% of general admission patients have physiological parameters outside the normal range within 24 hours prior to transfer to the intensive care unit (ICU). The lack of prompt and adequate response to patient deterioration may lead to an increase in their morbimortality and in health care costs [1].

Early identification of patients at risk of deterioration allows activation of specialized medical groups commonly called *Rapid Response Teams* that improve patient outcomes [2-5].

There are more than 100 detection systems for patients at risk, most of them derived from the original Early Warning Score (EWS). They can be divided into three groups according to the way the analysis is performed: simple parameter systems, multiple parameters or aggregate weighted systems. It is demonstrated that the latter are more accurate for the detection of cardiac arrest, mortality or transfer to ICU [6]. Particularly National Early Warning Score (NEWS) has a greater ability to discriminate patients at risk of cardiac arrest, unplanned admission to ICU, and death than the other 33 EWS in use [7].

The frequency of monitoring and recording novelties of a patient increases proportionally with their degree of deterioration [8]. Nursing has, therefore, a fundamental role in monitoring and documenting patients' behavior.

NEWS, like many early warning systems, is based on a simple measurement system on physiological parameters which are usually monitored in a hospital. It assigns scores to certain physiological variables depending on the degree of deviation from the normal range. The sum of the score of each variable results in the value of the score that allows to predict the risk of deterioration within the next 12 hours, depending on the result obtained [9].

When this indicator is used following a response protocol against each instance of the scale, it is able to prevent undesired clinical consequences. The parameters under consideration are as follows:

1. Respiratory rate
2. Oxygen saturation
3. Axillary temperature
4. Systolic Blood Pressure
5. Pulse frequency
6. Level of consciousness

NEWS makes an important contribution towards improving clinical care provided to inpatients by ensuring deteriorated patients are reviewed by the medical team with the appropriate level of training for each case, allowing the reduction of associated mortality. In one case, the effectiveness of reviewing important cases by trained physicians, alerted by nurses, increased by 87% [10].

The use of these patient assessment methods to predict risk is not widely distributed. In university hospitals with doctors and nurses being trained, their implementation is more likely because constant training of staff is necessary, although in some cases it is quite difficult [11].

The calculation of the score to predict patient deterioration depends on the existence of the six parameters; any missing value causes difficulties in generating alerts due to an error in such calculation [12, 13].

Sanatorio Finochietto is a private health institution, inaugurated in November 2013, with 125 adult general admission beds, 31 adult intensive care units, and 16 neonatal intensive care units. It is the first bio-eco-intelligent care center in Argentina, certified as a member of the global network of green and healthy hospitals. It has a commercial Health Information System (HIS) that allows electronic communication management between all the health team (e.g., doctors, kinesiologists, nurses, pharmacists, and nutritionists) and the administrative staff. It has interfaces with the effectors of complementary studies, such as laboratory studies, diagnostic imaging, and endoscopy. It uses mobile devices for safe medication administration and to record information at the foot of the patient's bed [14].

Although the measurements of the vital parameters associated with NEWS are carried out daily and stored in each patient's electronic medical record, calculation of such a score is not

simple since the measurements of some parameters are usually performed at different times.

The aim of this paper is to describe the procedure of score reconstruction by means of combining parameters of different times starting from the one with fewer measurements.

Methods

All patients transferred from general admission to the intensive care unit from 10 pm to 5 am from 14 November 2013 to 09 May 2016 were selected. The time selection necessarily represents patients who were suddenly decompensated during their stay in the general hospitalization area and an emergency transfer to the ICU was required to increase the level of complexity of their care. There are no scheduled patient movements in such time frame.

The period of data collection for each patient took place from the beginning of their hospitalization until the moment of their transfer to the intensive care unit in the selected time frame.

The parameters were obtained from computerized nursing records; five of the vital signs measurements (respiratory rate, oxygen saturation, axillary temperature, systolic blood pressure, and pulse rate) and the sixth, level of consciousness, were obtained from the structured assessment. All of them can be registered both from the office desktop computer and from the mobile application.

The state of consciousness is assessed by means of the Glasgow Coma Scale. Table 1 shows how these two scales are related.

Table 1– Glasgow Coma Scale adapted

NEWS State	Glasgow
Alert (A)	14-15
Voice (V)	11-13
Pain (P)	5-8
Unresponsive (U)	3

The scripts to process data were programmed in PERL as it is a general purpose language, which was originally developed for text manipulation, but is currently used for a wide range of tasks due to its ease of handling, its versatility and power [15].

Procedure description

This section presents the steps for reconstructing the score:

Step 1: Conduct the search for the parameters to be analyzed in the HIS database with the date of measurement and the pertaining patient.

Step 2: Store each parameter in a different table, keeping the date and the patient's ID.

Step 3: Determine the parameter with fewer measurements.

Step 4: Taking each of the values of the parameter having fewer measurements (in our case, the level of consciousness) as a reference, associate the data of each of the tables with each other, depending on the patient's ID and the measurement time, with a maximum of a one-hour difference. The one-hour margin is set for the data to have clinical significance.

Step 5: Eliminate the lines which are exactly the same to improve the performance of future processing of the information.

Step 6: Calculate the score value for each of the possible combinations of each set of 6 parameters of the same patient.

In Fig. 1 below, one can observe the problem and how the combination for the score calculation can be carried out. Given the same patient, called "A", in none of the four measurements (each row representing a measurement) are the six parameters complete. Each measurement is grouped according to the measured value for consciousness. Taking this last parameter as a starting point, the rest must be gone over making all the necessary combinations to complete the missing parameters for each of the measurements. In this way the number of scores per row will depend on the number of combinations that have been made to complete the missing parameters.

Step 7: Group the records for each patient.

Step 8: Given the existence of multiple measurements combined for a particular parameter of consciousness, keep the highest score and dismiss the rest of the records.

Step 9: Order the records based on the time when they were made. This will facilitate the analysis of the progression of patient deterioration.

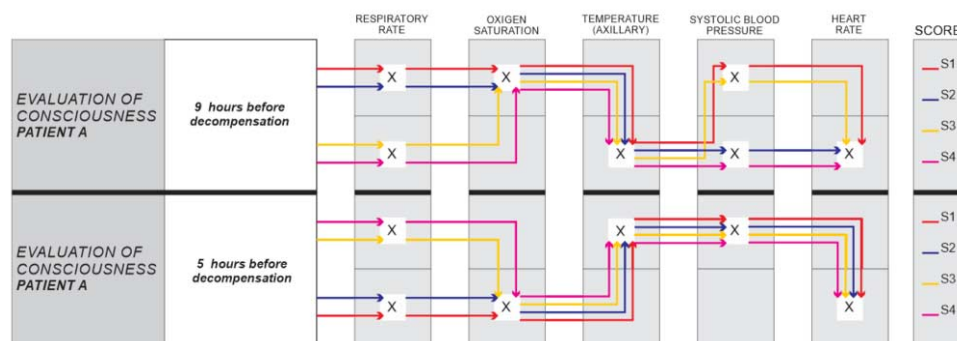


Figure 1– Reconstruction example

Concept Test

To illustrate the procedure, the case of patient 10793 is described. Initially, by means of the collection in the database, 4,045 combinations of parameters were obtained, based on

multiple assessments of consciousness, from the beginning of hospitalization to the nighttime transfer to the ICU. Two scripts were developed, the first one to solve step 6, and the other one, steps 7 to 9.

In the first instance, the repeated value filters were applied. This helped the information processing significantly due to the elimination of lines that were exactly the same. There were 730 different parameter combinations left. After this procedure, the first script for score calculation of each of the reconstructed combinations was executed.

Finally, the second script was executed to group all the measurements associated with a particular assessment of consciousness and to consider the highest result for each one. In addition, the script chronologically ordered the output information to facilitate the analysis of the progression of patient deterioration.

As a final result, seven score reconstructions were obtained for patient 10793, two of which were found within 12 hours prior to transfer to the ICU (Table 2).

Table 2 – Output example of algorithm

Patient	Hours prior to transfer to ICU	Score
10793	141	6
10793	96	17
10793	91	12
10793	48	7
10793	43	6
10793	11	5
10793	2	11

Results

The 283 patients analyzed generated 20,864,012 candidate lines to calculate the score from all possible parameter combinations, following the criteria of belonging to the same patient and being in more or less one-hour range focusing on the value of consciousness (limiting parameter). The processing of all the combinations generated by the second script delivered a total of 1,244 scores belonging to 110 patients. Finally, the scores obtained from the last 12 hours, prior to the transfer to the ICU, were filtered. The result obtained included 79 scores of 56 patients.

The manual process of collecting the parameters is conducted in two different screens of the system; precisely the parameter detected as limiting is that which is separated from the set. According to the data analyzed, the state of consciousness is first recorded and then the five remaining parameters (the physiological parameters) are recorded.

To represent the manual process, a search associating the physiological parameters with the state of consciousness within a range no longer than 10 minutes was conducted. Following the same criteria as those used for reconstructions, four measurements of four different patients were detected (Table 3).

Table 3 – Data recovery by method

Method	Patients	Scores
Manual process	4	4
Limiting parameter	56	79

The decision-making support system needs the existence of scores to generate the alarms. For the manual method, there would have been only four scores to analyze and generate the corresponding alerts. In contrast, the availability of scores increased substantially to 79 by the limiting parameter method. The greater availability of patient scores means that, given the same patient, there is more than one reconstruction

within 12 hours prior to the event. From the point of view of care, this means a great advantage since it allows to analyze the tendency towards deterioration in each patient with more than one reconstruction.

Some cases with very high scores were detected at moments far from transferring to the ICU that are related to a patient's previous stay in that area. There each patient is connected to a multi-parameter monitor that constantly sends the vital parameters to the HIS. This situation also caused the high number of lines initially found.

Finally, the four patients who had the records representing the manual process were compared with those obtained by means of the limiting parameter method. An exact match was found in three out of the four patients. The different value was of only 1 point; a 9 was obtained by means of the manual method and a 10 by means of the reconstruction by limiting parameter. This is within the expected range since the proposed method maintains only the highest possible value of the combination found (Table 4).

Table 4 – Score comparison

Patient	Manual process	Limiting parameter
22010	10	10
10793	9	10
24794	5	5
83004	3	3

Discussion

Decision support systems in the health field provide an objective and standardized basis for patient care. In particular, early deterioration detection systems can reduce risk of death or complications for the patient, and their associated care costs. NEWS is based on the assessment of six parameters for such deterioration prediction. The absence of at least one parameter of those necessary for the calculation makes it difficult to obtain the score, thus causing unnecessary delay in care that may lead to fatal consequences.

The number of full measurements made at the same time is scarce, but the results of the method proposed for the calculation of such measurements are promising as the coincidence is 75%. For the remaining value, although it was overestimated in the proposed method, it was the researchers' own decision how to handle the situation and its treatment context, given the sensitivity of the information.

Even though the patients considered from HIS are decompensated patients, in which case the score reconstruction in the previous hours is of utmost importance, another group of patients who decompensated during the day and was transferred to ICU were not taken into account.

Although the relation of the Glasgow scale to NEWS was arbitrary and there could be subtle variations, we believe that such variations should not affect the calculation, since the NEWS scale is much simpler as to the state of consciousness.

This study is expected to continue with the validation of the reconstruction procedure proposed against the manual process of measuring the six parameters for the same moment in a large sample of patients. The calculation will be made according to the method described at a certain time of the day. On the same patients, a trained nurse, without knowing the existing values, will carry out the measurements and record the six parameters for each patient. Finally, the score will be calculated for both methods and their differences will be analyzed.

Finally, we intend to determine the optimal relation between the time difference existing among the parameters and the degree of accuracy of the reconstructed score. For such a purpose, we will reconstruct scores with values that are increasingly more remote from the limiting parameter. We will start from reconstructions of up to 60, 120, 180, 240 and 360 minutes, which will be compared with those scores reconstructed with a 5-minute maximum difference from the limiting parameter. This analysis will allow us to make the maximum time value from which we can move away from the limiting parameter precise, without the score losing accuracy.

Conclusion

The information treatment procedure described in this paper enables the improvement of the information availability in more than 10 times with respect to the number of patients and almost 20 times for the amount of such patients' scores, taking into account the current registration ways. This information will reduce the risk of serious events in hospitalized patients via early intervention, despite not initially having all the full values.

Acknowledgements

First of all, we would like to thank Dr. Jorge Kritzer for being the source of inspiration on the subject. We would like to make special mention of Dr. Ramón García Martínez for having supported us by devoting his time to correct and guide us in the analysis and presentation of our research.

In memoriam to Dr. Ramón García Martínez.

References

- [1] L. Tarassenko, A. Hann, and D. Young, Integrated monitoring and analysis for early warning of patient deterioration, *British Journal of Anaesthesia* **97** (2006), 64-68.
- [2] M.A. Peberdy, M. Cretikos, B.S. Abella, M. DeVita, D. Goldhill, W. Kloeck, S.L. Kronick, L.J. Morrison, V.M. Nadkarni, G. Nichol, J.P. Nolan, M. Parr, J. Tibballs, E.W. van der Jagt, L. Young, R. International Liaison Committee on, A. American Heart, C. Australian Resuscitation, C. European Resuscitation, Heart, C. Stroke Foundation of, F. InterAmerican Heart, A. Resuscitation Council of Southern, C. New Zealand Resuscitation, C. American Heart Association Emergency Cardiovascular Care, P. American Heart Association Council on Cardiopulmonary, C. Critical, C. Interdisciplinary Working Group on Quality of, and R. Outcomes, Recommended guidelines for monitoring, reporting, and conducting research on medical emergency team, outreach, and rapid response systems: an Utstein-style scientific statement: a scientific statement from the International Liaison Committee on Resuscitation (American Heart Association, Australian Resuscitation Council, European Resuscitation Council, Heart and Stroke Foundation of Canada, Inter-American Heart Foundation, Resuscitation Council of Southern Africa, and the New Zealand Resuscitation Council); the American Heart Association Emergency Cardiovascular Care Committee; the Council on Cardiopulmonary, Perioperative, and Critical Care; and the Interdisciplinary Working Group on Quality of Care and Outcomes Research, *Circulation* **116** (2007), 2481-2500.
- [3] M. Capan, J.S. Ivy, T. Rohleder, J. Hickman, and J.M. Huddleston, Individualizing and optimizing the use of early warning scores in acute medical care for deteriorating hospitalized patients, *Resuscitation* **93** (2015), 107-112.
- [4] U. Kyriacos, J. Jelsma, M. James, and S. Jordan, Monitoring vital signs: development of a modified early warning scoring (MEWS) system for general wards in a developing country, *PloS One* **9** (2014), e87073.
- [5] S. Romero-Brufau, J.M. Huddleston, J.M. Naessens, M.G. Johnson, J. Hickman, B.W. Morlan, J.B. Jensen, S.M. Caples, J.L. Elmer, J.A. Schmidt, T.I. Morgenthaler, and P.J. Santrach, Widely used track and trigger scores: are they ready for automation in practice?, *Resuscitation* **85** (2014), 549-552.
- [6] M.M. Churpek, T.C. Yuen, and D.P. Edelson, Risk stratification of hospitalized patients on the wards, *Chest* **143** (2013), 1758-1765.
- [7] G.B. Smith, D.R. Prytherch, P. Meredith, P.E. Schmidt, and P.I. Featherstone, The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death, *Resuscitation* **84** (2013), 465-470.
- [8] S.A. Collins, K. Cato, D. Albers, K. Scott, P.D. Stetson, S. Bakken, and D.K. Vawdrey, Relationship between nursing documentation and patients' mortality, *American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses* **22** (2013), 306-313.
- [9] R.C.o.P.o. London, *National Early Warning Score (NEWS): standardising the assessment of acute-illness severity in the NHS*, Royal College of Physicians, London, 2012.
- [10] L. Spiers, J. Singh Mohal, J. Pearson-Stuttard, H. Greenlee, J. Carmichael, and R. Busher, Recognition of the deteriorating patient, *BMJ quality improvement reports* **4** (2015).
- [11] M. Niegsch, M.L. Fabritius, and J. Anhøj, Imperfect implementation of an early warning scoring system in a Danish teaching hospital: a cross-sectional study, *PloS One* **8** (2013), e70068.
- [12] D.A. Clifton, L. Clifton, D.-M. Sandu, G.B. Smith, L. Tarassenko, S.A. Vollam, and P.J. Watkinson, 'Errors' and omissions in paper-based early warning scores: the association with changes in vital signs—a database analysis, *BMJ open* **5** (2015), e007376.
- [13] C.F. Gordon and D.J. Beckett, Significant deficiencies in the overnight use of a Standardised Early Warning Scoring system in a teaching hospital, *Scottish Medical Journal* **56** (2011), 15-18.
- [14] H. Navas, L. Graffi Moltrasio, F. Ares, G. Strumia, E. Dourado, and M. Alvarez, Using mobile devices to improve the safety of medication administration processes, *Stud Health Technol Inform* **216** (2015), 903.
- [15] The Perl Programming Language - www.perl.org, in.

Address for correspondence

Dr. Hernán Navas
Chief of Health Informatics. Sanatorio Finochietto, Av. Córdoba
2678, CP C1187AAN, Ciudad Autónoma de Buenos Aires.
hnavas@sanatoriofinochietto.com

Framework of Performance Measures for Health Information Exchange (HIE)

Okada Mihoko ^a, Ishida Haku ^b, Watanabe Sunao ^c, Shiratori Yoshimune ^d

^a Department of Health Informatics, Kawasaki University of Medical Welfare, Kurashiki City, Okayama, Japan

^b Department of Medical Informatics & Decision Sciences, Yamaguchi University Hospital, Ube City, Yamaguchi, Japan

^c St. Luke's MediLocus, St. Luke's International Hospital Brach Clinic, Chiyoda-ku, Tokyo, Japan

^d Medical IT Center, Nagoya University, Nagoya City, Aichi, Japan

Abstract

Health Information Exchange (HIE) is the most prevalent patient information sharing technology currently in use. Although a number of regional healthcare information systems exist in Japan, little is known about the effectiveness of the systems or how effectiveness should be measured. As an approach to this issue, we developed a framework for measuring effectiveness of HIE using the dimensions for health indicators from ISO/TS 21667:2004 "Health Indicators Conceptual Framework". Three phases for measuring HIE are defined: Phase I: the static aspects of a system; Phase II: the use of the system; Phase III: the outcomes of the use of the system. Complex factors involved in HIE are organized and objectives of evaluation are made clear. The domains to which measures are applied and categories of measures are defined. Sample measures extracted from publications and co-authors' studies are discussed. This work is the first step towards the systematic development of a framework of measures of the effectiveness of HIE.

Keywords:

Health Information Exchange, Quality Indicators, Cost Effectiveness, Information Systems

Introduction

It is expected that health information technology can facilitate information sharing regionally, nationally, and internationally to facilitate continuity and quality of care. In recent years, the term "HIE (Health Information Exchange)" has become the most widely used term for regional healthcare information systems.

According to the web site of the US government HealthIT.gov, HIE is defined as follows:

HIE generally refers to sharing of clinical data between health care institutions for efficiency, cost-effectiveness, quality care, and for patient safety.

A number of regional healthcare information systems exist in Japan, where hospitals, clinics and pharmacies share patient information electronically. According to a working paper published by Japan Medical Association Research Institute in 2015, more than 200 systems exist in Japan [1].

Systematic reviews and a number of original papers that focus on effectiveness of HIE discuss major concerns about regional healthcare information systems' effectiveness [2-8]. Even though a huge amount of money has been expended, systematic effectiveness analyses of these systems has seldom been conducted.

Effectiveness of a healthcare information systems is difficult to evaluate because multiple factors are involved including human (personnel), organization, workflow/process, as discussed by Sittig and Singh in their paper "A New Socio-technical Model

for Studying Health Information Technology in Complex Adaptive Healthcare Systems" [9].

To systemically assess the effectiveness of regional healthcare information systems, we investigated what and how measures were obtained, what measures are feasible, how the various factors may be captured and organized, and how efficiency and effectiveness can be evaluated.

Our study focuses on effectiveness of regional healthcare information systems where patient information is shared between hospitals, clinics and pharmacies. Since the objectives and the concepts of the systems are represented by HIE, we will use the term HIE and regional healthcare information systems interchangeably.

As a result of our study, we have developed a framework for measures of HIE. We present the major results and discuss issues that should be addressed.

Methods

The methods and the process for the Framework development are shown below (Figure 1).

1. International Standards and Technical Reports on health systems performance indicators that discuss both qualitative and quantitative aspects were examined [10], and the dimensions for measures of HIE were derived.
2. Domains to be measured and categories of measures were defined based on published papers on HIE [2-8].
3. The domains to be measured and categories of measures are reviewed in the light of the Donabedian model that provides a framework for examining health services and evaluating quality of healthcare.
4. A survey on regional healthcare institutions and users (clinicians) in Japan were conducted and the results were reflected in the framework. Some measures were investigated for feasibility and usefulness.
5. Co-authors conducted several studies on some measures experimentally, and the results were reflected in the framework. Quality indicators adopted in the co-authors' institutions were also investigated.

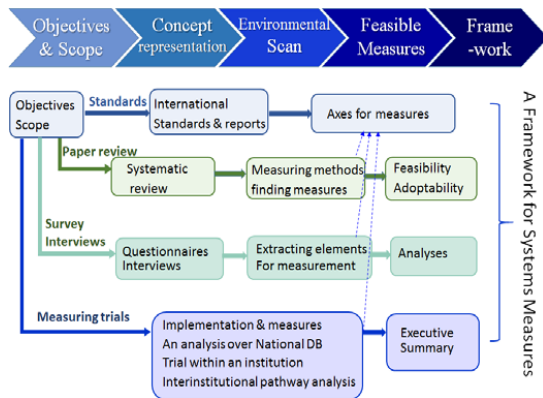


Figure 1 - Framework development: the methods and process

Results

Conceptual Representation of Healthcare and HIT

When improvements in continuity of care are observed after adoption of hospital information systems (HISs), it may be due to changes in human attitudes, workflow, care process, etc. Sit-tig and Singh presented a model with eight dimensions in their article “Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems,” stating that HIT interventions must be understood in the context of their simultaneous effects across multiple dimensions of the model [9].

We present a conceptual representation of healthcare and HIT in Figure 2. The areas in orange show healthcare itself regardless of existence of HIT. Concepts such as “quality healthcare”, “quality measures”, “patient engagement” exist regardless of HIT. The areas in blue are components of HIT such as EHR or “data interoperability.” The overlaps between the concepts in orange and blue areas are apparent. HIT is not a simple combination of healthcare and IT. It is a new healthcare domain established with the advent of Health Informatics.

Effectiveness Measures - Literature Review

We have investigated papers on effectiveness of HIE. There are a number of original papers and several reviews including, “Outcomes from Health Information Exchange: Systematic Review and Future Research Needs” by Hersh, et al [3]. It is an in-depth review discussing the limitations of the studies under review. In their review, 34 studies on outcomes of HIE were identified. Most of the papers on HIE effectiveness focused on healthcare resources. Some papers discussed a decrease in laboratory testing and amount of costs. For hospital admissions, some show a decrease in the number of admissions, but others show no decrease. HIE in general reduces duplicative laboratory and radiology testing, emergency department costs, and hospital admissions. There are also reports on public health reporting and ambulatory quality of care. But papers on the effects of HIE on clinical outcomes have not been found.

Framework of Effectiveness Measures of HIE

Eight Dimensions of Measures

International organizations such as ISO and OECD published frameworks for health quality or health systems performance. For example, “ISO/TS 21667:2004 Health informatics - Health Indicators Conceptual Framework” shows performance indicators for health systems with eight dimensions below [10]:

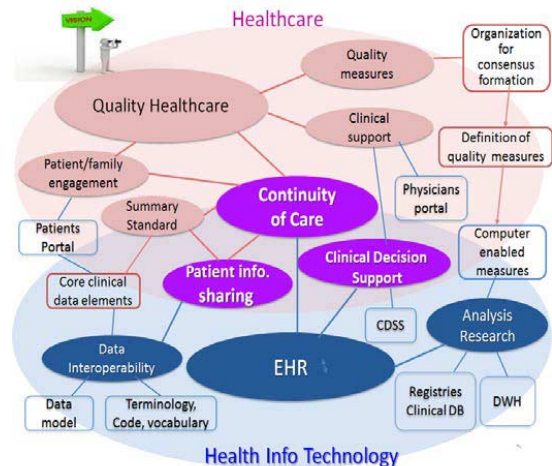


Figure 2 - Healthcare and Health Information Technology

- **Acceptability:** All care/services provided meets the expectations of the client, community, providers and paying organizations, recognizing that there may be conflicting, competing interests between stakeholders, and that the needs of the clients/patients are paramount.
- **Accessibility:** The ability of clients/patients to obtain care/service at the right place and the right time, based on respective needs.
- **Appropriateness:** Care/service provided is relevant to the clients'/patients' needs and based on established standards.
- **Continuity:** The ability to provide uninterrupted coordinated care/service across programmes, practitioners, organizations, and levels of care/service over time.
- **Competence:** An individual's knowledge and skills are appropriate to the care/service being provided.
- **Effectiveness:** The care/service, intervention or action achieves the desired results.
- **Efficiency:** Achieving the desired results with the most cost-effective use of resources.
- **Safety:** Potential risks of an intervention or the environment are avoided or minimized

The eight dimensions of performance indicators are for health system performance. We applied the eight dimensions above to health information systems or HIE in our framework, using the term “measures” instead of “indicators”.

Effectiveness Measures – Three Phases

Various aspects are involved in evaluating HIE, including organizations, personnel, operations, system functionality, system use rate, patient participation, provider satisfaction, and patient outcomes. Taking these into account, we arranged performance measures of HIE in the three phases (Figure 3):

Phase I: System organization (including both IT and human)

Phase II: Use of HIE

Phase III: Effects by the use of HIE

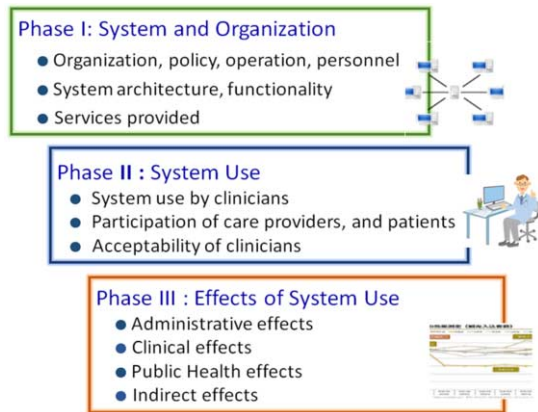


Figure 3 - Measuring HIE: Three Phases

It does not necessarily mean that the HIE should be measured in this order. Phase I may include such domains to be measured as organization, policy, operation, system functionality and services provided. Phase II includes domains such as use by clinicians, participation of patients, acceptability by healthcare providers, and patient satisfaction. Phase III includes domains such as administrative effects and clinical effects. There may also be public health effects, and indirect effects

such as conformance rate of clinical guidelines.

We investigated how measures may be captured from the view point of Donabedian's "Structure, Process, and Outcome" [11]. We considered HIE not only as an information technological system, but also as a socio-technological system that includes organizations, personnel, operation and so on. From this point

of view, we considered the Phase I "Structure", and Phase II "Process." Phase III was considered "Outcome." For Phase III, we represented outcomes if they measure aspects resulting (directly or indirectly) from the use of the system. Hence the measures in Phase III may be classified as "Process" rather than "Outcome" from the view point of clinical quality measures.

Overview of the Framework

Figure 4 shows the overview of the developed framework for HIE effectiveness measures. Shown in the columns are eight dimensions for performance: "Acceptability, Accessibility, Appropriateness, Competence, Effectiveness, Efficiency, Safety, and Continuity." In the rows, the major domains are organized into the three phases. Then in each domain, categories of measures are shown.

Example measures

The framework does not include concrete measures, but some measures are shown as examples in Figure 5. Measures for Phase I and Phase II are mostly straightforward. Some of them are taken from our survey carried out in 2014 in Japan across five regional healthcare information systems. For a given measure, the relevant dimensions are shown by a check mark. A large thick check mark indicates strong relevance, while a smaller check mark indicates weak relevance. For example, for "the number of years in operation" in the domain "information system organization," a large check mark is shown under the column "Continuity."

Most of the sample measures in Phase III were found in the literature. We reviewed the papers and extracted measures that are generally considered feasible. Papers on HIE effectiveness mostly focus on healthcare resources. For example, "amount of reduced testing/imaging" in the domain "Laboratory & other tests" is found frequently in literature review, and is most relevant to the "Efficiency" dimension and also relevant to "Appropriateness", "Continuity" and "Safety".

Phases of Measuring HIE	Measured Domains	Category of Measures	Example Measures	Acceptability	Accessibility	Appropriateness	Competence	Continuity	Effectiveness	Efficiency	Safety
Phase I	System Organization	Information system organization									
		Management/Staff organization									
		Operation policy									
System construct Organization	System Functionality, Services	User interface									
		Support for patient engagement									
		Data interoperability									
Phase II	Use of System	Support for secondary use									
		Services to care providers									
		Institution participation									
Phase III	Effects of System Use	Patient participation									
		Use by care givers									
		Care givers satisfaction									
Effects of System Use	Administrative, Clinical effects	Cost per participation									
		Laboratory & other tests									
		Admission									
Effects of System Use	Public Health effects	Emergency department visit									
		Drug administration									
		Preventive medicine									
Effects of System Use	Other/indirect effects	Quality measures for a given disease									
		Information sharing among relevant institutions									
		Clinical guideline compliance rate									

Figure 4 - Overview of the developed Framework of Measures for HIE

Some measures are from the co-authors' work or used in their institutions. For example, "improvement of HbA1c (Surrogate)" and "cancer survival rate (True)" are measures for the domain "quality measures for a given disease," and a check mark is shown in the "Effectiveness" column. Measures such as "rate of BP lower than 140/90" is one of compliance rates of a clinical guideline under the domain "Other/indirect effects."

measures are mostly descriptive statistics, and although there may be discussions about what measures should be adopted, measuring should not be difficult. Many of the sample measures shown are taken from our survey on five regional healthcare information systems. For example, "How much would you spend for the system you participate in?" was a question for clinicians; we found the median acceptable cost was almost equal among five systems.

In the past, there were no clear guidelines for measures, and the

Domains	Category of Measures	Example Measures	Acceptability	Accessibility	Appropriateness	Competence	Continuity	Effectiveness	Efficiency	Safety
System Organization	Information system organization	number of years in operation initial/maintenance costs					✓		✓	
	Governance/Staff	staff (engineers, administrative, etc.)			✓					
	Operation policy	business continuity planning patient consents(including withdrawal)			✓					
					✓					
System Functionality, Services	Support for patient	information provided to patients			✓					
	Data interoperability	shareable (exchangeable) information standards adopted (terminology, code) bidirectional exchange				✓				
						✓				
	Data storage and usability	disaster countermeasures usability for research					✓			✓
	Services to clinicians	clinical measures provided to clinicians				✓			✓	
Use of System	Institution participation	rate of participating clinical institutions (hospitals, clinics, pharmacies)			✓					
	Patient participation	rate of registered patients patient satisfaction	✓	✓						
	Use by care givers	rate of care givers using the system frequency of the use of the system	✓	✓						
	Care givers satisfaction	how much is worth spending	✓						✓	
	Cost per participation	cost per size of the system							✓	
Administrative, Clinical effects	Laboratory & other tests	amount of reduced testing/imaging cost savings due to reduced testing			✓		✓		✓	✓
	Admission	reducing hospital admissions lower costs for (re)admissions							✓	
	Emergency dept. visit	reducing emergency dept. revisits				✓			✓	
	Drug administration	decrease in duplicated of medication decrease in inadequate prescription by medication reconciliation					✓	✓	✓	✓
	Preventive medicine	increase in vaccination rate, e.g. pneumococcal vaccination			✓		✓			
	Quality measures for a given disease	improvement of HbA1c (Surrogate) cancer survival rate(True)						✓		
Public Health effects	Information sharing among relevant institutions	improved completeness of reporting for notifiable diseases improved identification of HIV patients for follow-up care clinical guidelines compliance rate,e.g.,			✓	✓				
	Other/indirect effects	rate of HbA1c<7.0% rage of BP lower than 140/90						✓	✓	

Figure 5 - Framework of Measures for HIE with sample measures

Discussion

Implementation of HIE measures of Phase I and Phase II

The term and the concept of HIE is in use more and more in the field of Health Informatics. HIE mostly refers to the electronic information sharing among clinical institutions, for efficiency, cost-effectiveness, quality care and safety.

In the framework presented, three phases are defined for measuring the effectiveness of HIE. For Phase I and Phase II,

framework should serve as a sharable tool for governance of the regional healthcare information systems.

Implementation of HIE measures of Phase III

For measures of the third phase, i.e., measures for outcomes, further discussion may be necessary. Among a number of published papers, the systematic review by Hersh et al, which outlines the need for more research, aligns closely with our current work. They discussed the significant limitations of the evidence base, and showed four primary limitations of the available evidence on the impact of HIE (and Health IT in general). That is,

(1) suitability of study design; (2) execution of the studies; (3) complexity of the interventions with implications for interpretation and for generalizability; and (4) changes in the technology or policy governing its use. The evidence level of the studies investigated was considered low to moderate due to multiple limitations [3].

A primary limitation is due to the complexity of interventions. The HIE itself is necessarily only part of a more complex intervention, and the mixed effects might change the behavior of clinicians or others in the health systems. For example, if HIE effect on the compliance rate of electronic clinical guidelines is measured, multiple interventions (factors) are involved. When some sample measures are applied in clinical settings that employ a HIE, possible factors involved other than HIE may need to be considered when reporting the results.

Framework as an approach to Health IT evaluation

The US Agency for Healthcare Research and Quality published the report "Identification and Prioritization of Health IT Patient Safety Measures, Final Report" [12], which discusses that while the use of Health Information Technology (HIT) presents many new opportunities to improve patient care and safety, the complex interactions between people, processes, culture, and technology can also create an environment where new hazards are introduced. There is a need for measures to help identify the nature, scope, and prevalence of HIT-related safety issues and to assess how well providers, vendors, and others are preventing and/or mitigating HIT-related safety concerns.

Similar discussions apply to the third phase of measures of the framework developed in the present study. The concepts and sample measures for "Outcome" serve as the groundwork for the complex discussions of effectiveness of regional healthcare information systems.

Conclusion

Healthcare information systems and HIT are widely viewed as essential to the transformation of healthcare to counteract rising costs, inefficiency, preventable errors, and quality of care. HIT is expected to yield benefits in quality, safety, and efficiency of healthcare. The Japanese government as well as the US and other countries' governments have been engaged in a concerted effort to promote adoption of HIT. Despite the considerable investment and the widespread rapid adoption of HIT, little is known about the effectiveness or cost-effectiveness of the systems, that is, how useful the systems are.

As an approach to this issue, we developed a framework for measuring the effectiveness of HIE. The framework comprises of three phases of measures and eight dimensions of quality. It is recommended that responsible organizations measure their systems for monitoring and for governance at Phase I and II. For Phase III, due to complex factors involved and limitations of studies, literature reviews showed that the evidence level is relatively weak. This is generally true when an outcome measure is applied to a specific setting, and considerations may be necessary in presenting the results as the effects may be the mix of HIE and other factors.

We conclude that the framework will assist administrators, healthcare providers, researchers and regional communities as a tool to measure the effectiveness of regional healthcare information systems, thereby contributing to the evaluation and improvement of their systems, and ultimately improve quality care.

Acknowledgements

The study was funded during 2015 - 16 by Health and Labour Sciences Research Grants, the Ministry of Health, Labour and Welfare.

References

- [1] Overview of IT-based regional healthcare collaborations, 2015, Japan Medical Association Research Institute, available at <http://www.jmari.med.or.jp/download/WP368.pdf> (original in Japanese).
- [2] Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med*. 2006; 144 (10): 742.
- [3] Hersh WR, Totten AM, Eden KB, Devine B, Gorman P, Kassakian SZ, et al. Outcomes From Health Information Exchange: Systematic Review and Future Research Needs. *JMIR Med Inform*. 2015; 3 (4), e39: 1.
- [4] Kern LM, Ancker JS, Patel EA, Dhopeswarkar RV, and Kaushal R. Evaluating health information technology in community-based settings: Lessons learned. *J Am Med Inform Assoc*. 2011; 18(6): 749.
- [5] Ofir BA, Shabtai I and Leshno M. The impact of EHR and HIE on reducing avoidable admissions: Controlling main differential diagnoses. *BMC Med Inform Decis Mak*. 2013; 13.
- [6] Ross SE, Radcliff TA, Leblanc WG, Dickinson LM, Libby AM, and Nease DE Jr., Effects of health information exchange adoption on ambulatory testing rates. *J Am Med Inform Assoc*. 2013; 20(6): 1137-1142.
- [7] Bailey JE, Wan JY, Mabry LM, Landy SH, Pope RA, Waters TM, and Frisse ME, Does health information exchange reduce unnecessary neuroimaging and improve quality of headache care in the emergency department? *J Gen Intern Med*. 2013; 28(2): 176-183.
- [8] Vest JR, Kern LM, Silver MD, Kaushal R. HITEC investigators. The potential for community-based health information exchange systems to reduce hospital readmissions. *J Am Med Inform Assoc*. 2015; 22(2): 435-442.
- [9] Sittig DF and Singh H. A New Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems, *Qual Saf Health Care* 2010; (Suppl 3): i68.
- [10] ISO/TS 21667:2004 Health informatics - Health Indicators Conceptual Framework, 2004.
- [11] Donabedian A. Special article: The quality of care: How can it be assessed? *Archives of Pathology & Laboratory Medicine*; Nov 1997; 121, 11; ProQuest Nursing Journals. Reprinted with permission from *JAMA* September 23/30, 1988; 260 (12):1743. Copyright 1988, American Medical Association.
- [12] Agency for Healthcare Research and Quality: Identification and Prioritization of Health IT Patient Safety Measures, Final Report, February 11, 2016.

Corresponding author:

Mihoko Okada

Email: m-okada@mxq.mesh.ne.jp

Current Affiliation: Translational Research Informatics Center, 1-5-4 Minato-Jima, Minami-Machi, Chuo-Ku, Kobe City, 650-0047 Japan

A Comparison of Two Principal Systems for Monitoring of Technology-Induced Errors in Electronic Health Records

Sari Palojoki^a, Elizabeth M. Borycki^b, Andre W. Kushniruk^b, Kaija Saranto^a

^a University of Eastern Finland, Department of Health and Social Management, Finland

^b School of Health Information Science, University of Victoria, Victoria, Canada

Abstract

Current methods for monitoring harm caused by health information technology (HIT) are minimal, even if there are known risks associated with the use of HIT. Monitoring is predominantly based on voluntary reporting using generic patient safety adverse events reporting systems. Another important means for monitoring technology-induced errors is a health authority reporting system. International oversight systems have medical devices' related software's adverse event and failure reporting models, but these systems differ due to differences in the legislation. The protocol for this study included an electronic database literature search and the eliciting of information for study purposes from the literature. The purpose is to provide a scoping review focused on two types of systems and provide implications for monitoring technology-induced errors in the future. The analysis revealed not only differences, but also similarities between these systems which raises the question of these systems' effectiveness due to overlapping goals in collecting data.

Keywords:

Electronic Health Records; Patient Safety, Medical Informatics

Introduction

A recent systematic review of eHealth technologies and their impact on the safety of healthcare showed that the problem of electronic health record (EHR) safety is an existing and possibly critical issue [1]. It is assumed that an increase in the implementation of information technology within healthcare systems will potentially lead to patient safety incidents by introducing novel vulnerabilities [2]. In 2012, the Institute of Medicine (IOM) [3] recommended that patient safety incidents which relate to the use of Health Information Technology (HIT) should be monitored and that this information be used to improve the safety of HIT. By detecting these errors, there is an opportunity to decrease risk for patients [4].

HIT related medical errors are defined in several ways. The concept of technology-induced error as defined by Borycki [4] has increasingly appeared in the literature during recent years. These errors result from 'the design and development of technology, the implementation and customization of a technology, and the interplay between the operation of a technology and the new work processes that arise from a use of technology' [4].

Existing methods for monitoring adverse events and harm caused by HIT systems rely mostly on voluntary reporting using generic patient safety adverse event reporting systems [5-6]. These adverse event monitoring systems have multiple functions, including: (a) the monitoring of levels of harm, and

(b) identifying rare events and disseminating knowledge about patient safety issues arising during healthcare to avoid these incidents recurring [7]. These systems form a mechanism which also includes the analysis and triggering of an investigation of an incident [8] whether on a mandatory or voluntary basis. It is believed that voluntary adverse event monitoring systems are the most effective way of avoiding liability issues for providers, but there is also the belief that there is a need for greater accountability. Incident reporting systems generate numerators without denominators which has been regarded as a major problem [9-12].

HIT has not been strictly regulated over the past few decades [13]. Today, international healthcare systems have medical device related software adverse event and failure reporting models, but there are differences between these systems [13-18]. The differences originate from different legislation. In the United States (US), regulatory requirements that can be used to evaluate EHR system safety are insufficient [16-18]. Software has been regarded as a medical device according to European Union (EU) legislation for almost a decade. Medical device regulations apply to medical software in the EU [18]. The EU directive focuses on pre-market testing aiming for European conformity. In general, the regulatory oversight systems seek to gather data to help HIT developers and clinicians better understand and mitigate risks associated with HIT implementation and use [19, 21].

In this paper the authors provide: (a) review two of the most dominant reporting system types, (b) outline the implications of monitoring technology-induced errors in HIT and (c) provide insights into future research directions based on these analyses. More specifically, the authors answer the following research questions: (1) What are the main characteristics of the two dominant HIT related monitoring systems, (2) What are differences between these systems? and (3) What are the implications of this research for future analyses?

The scoping review presented in this paper is limited to the perspectives of healthcare organisations' (i.e., the HIT vendors view is excluded). For the purpose of this paper the term patient safety incident reporting system is referred to as "IRS". Different types of software's regulatory authority adverse event and failure reporting models and vigilance systems are referred to as "oversight systems".

Methods

A scoping review was conducted by the authors using Arksey's and O'Malley's approach [20]. The following steps were undertaken: (1) an electronic database was searched, and (2) information was elicited from downloaded articles to answer study questions. The database search was performed on

PubMed. Search results were restricted to those references published in English only. The search was performed on the two main topics associated with the theme of this work: HIT and monitoring. A combination of keywords were used: "HIT" OR "EHR", OR "medical software" and "monitoring" OR "patient safety reporting" OR "incident reporting system" OR "surveillance", "regulation" OR "oversight" OR "vigilance" (Moreover, an additional search was based on these previous results which were combined with a keyword "technology induced error", and this resulted in 3 citations of which one was a duplicate.)

Initially, the researchers reviewed the abstracts of the downloaded publications. Those abstracts that fulfilled the following criteria: patient safety incident reporting system and an authority surveillance perspective were downloaded and reviewed further. An additional type of non-scientific material collected contained legislation documents. Finally, the following data were extracted by two reviewers from the articles using the following criteria: the main objectives of reporting systems, nature, confidentiality aspects and reporting modes, who are the reporters, what is reported, and analysis of the incidents. The findings arising from our work are reported according to these data extraction criteria.

Results

In the next section of this paper, we report on our findings arising from the scoping review in the following areas: main objectives of reporting systems, nature, confidentiality aspects and reporting modes, who are the reporters, what is reported, and analysis of the incidents. Figure 1 illustrates the detailed steps of the literature search in PubMed. A total of 4966 citations were received of which 18 articles were included in the final review. We present our findings in terms of the two types of reporting systems described in the literature: oversight (ORS) and incident reporting systems (IRS).

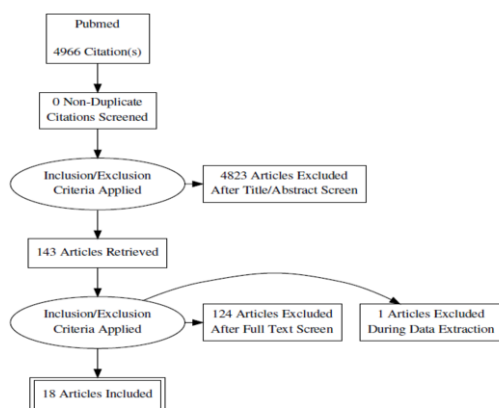


Figure 1. A PRISMA flow diagram of the review.

Main Objectives of Reporting Systems

Oversight system

The primary purpose of the EU Medical Device Vigilance System, which is grounded in the Medical Device Directive 93/42/EC, is to "improve the protection of health and safety of patients, users and others by reducing the likelihood of reoccurrence of the incident elsewhere". The Medical Device Vigilance System aims to facilitate an early and harmonised implementation of 'Field Safety Corrective Action' across the

Member States where the device is in use, in contrast to action taken on a country by country basis [18, 20].

Existing FDA databases for medical device errors are focused on collecting data about medical devices and have only recently been used to collect some data on reporting EHR-related incidents. The Office of the National Coordinator (ONC) has recently created a HIT complaint website; however, the form for entering complaints is basic with few specific HIT and error questions to answer limiting the usefulness of the resulting database. The regulatory oversight system's intention is to gather data to help HIT developers and clinicians better understand and mitigate risks associated with HIT implementation and use [8, 15, 19].

Incident reporting system

The aims of a patient safety incident systems (IRS) are broad. IRS systems are intended to monitor levels of harm, identify rare events and disseminate knowledge about patient safety of care. Learning from errors is the main goal of IRS [7-8, 22].

Legislation

Oversight system

Directive 2007/47/EC1 amended the definition of the term "medical device" used in Directive 93/42/EEC, and stand alone software with medical purpose are subject to medical device directives [18].

The FDA currently considers clinical information systems to be medical devices, but to date their regulatory requirements are not enforced. The Office of the National Coordinator (ONC) is collecting data and attempting to resolve issues that have been submitted to their complaint system. In the US, no government agency is currently fully equipped to perform regulatory and legal authority functions where Medical software is concerned [15-17, 19].

Incident reporting system

EU Council Recommendation 2009/C151/012 [23] regarding reporting and learning about incidents recommends that Member States support the establishment or strengthen blame-free reporting and learning systems regarding adverse events, which provide information about the extent, types and causes of errors, adverse events and near misses. In the US the Patient Safety and Quality Improvement Act was launched in 2005. The core goals of the act are to encourage health care professionals to improve the safety of health care, to understand the underlying causes of hazards, and to share the results, thereby minimizing risks related to patient care [27].

Nature, Confidentiality Aspects and Reporting Modes

Oversight system

ORS in the EU are mandatory and failing to report an event may be punishable [e.g., 26]. The US databases supported by the FDA: Medical Device Reporting (MDR) cover mandatory reporting from 1984-1996 and voluntary reporting thru to June 1993. Manufacturer and User Facility Device Experience (MAUDE) contains voluntary reports starting in June 1993, and facility reports starting in 1991 [15]. The ONC Health IT Complaint system is newly active and posts information about how to address issues [20]. The aspect of confidentiality does not apply to vigilance systems, when it comes to the principle of anonymity, and as is the case in the voluntary based IRSs, which are described in the next section of this paper. Vigilance systems are based on the traceability and accountability of the event and confidentiality would limit both traceability and accountability [19, 21, 25]. Paper-based reporting systems as well as electronic formulas are used, e.g., in Finland both are in

use, with the data structure of the report always remaining the same [26].

Incident reporting system

IRS reporting systems often contain anonymous data from both mandatory and voluntary systems. Closely connected to the issue of mandatory or voluntary reporting are the questions of reporting confidentiality. The discussion is focused around the aspect of preserving the reporter's anonymity. If healthcare personnel perceive that they will suffer judicial or legal consequences, when reporting patient safety incidents, they are less likely to report an incident [27]. To date the research suggests that the use voluntary IRSs continues to be the most effective way of encouraging reporting in a nonpunitive or "no blame" culture. The opposite view also exists, e.g., the public feels that mandatory reporting improves accountability [10-11].

There are big differences between these reporting systems in EU Member States. In the EU both mandatory and voluntary incident reporting systems exist across states; for example, in Finland, hospitals are required to have reporting systems (i.e., mandatory), but the reporting of adverse events (i.e., errors) by health professionals is a voluntary activity. In Finland, there is also a focus on reporting on near misses. In the Netherlands, healthcare professionals are obliged to report serious incidents to the Health Care Inspectorate. Yet, the reporting of incidents by health professionals is voluntary and recommended by professional organisations (EU). Three different types of national patient safety incident reporting systems are used to collect adverse event data: systems for sentinel events only, systems focusing on specific clinical domains (e.g. intensive care, emergency room) and healthcare system-wide, comprehensive reporting systems [27].

Traditional paper-based incident reporting systems as well as new forms of IT are used to enhance reporting. It has been suggested that electronic health records could support new applications such as surveillance of patient safety events e.g., the integration of an IRS into an EHR used in operating theaters that has been implemented to allow for the reporting of accidents and preventable complications [28].

Who Are the Reporters?

Oversight System

Research suggests that health professionals need to be involved in reporting. For example, in the EU for a monitoring systems to be effective, user involvement is regarded important. There are differences between EU member states in the EHR user involvement [19, 29]. For instance, in Finland there is an obligation for the professional user of Medical software to report HIT related safety flaws but in most EU countries the user reporting is not mandatory [26].

Incident Reporting System

Reporting is typically done by frontline personnel where IRS are used. In recent years, patient involvement in reporting has increased, even though this involvement is voluntary in nature for organizations [27].

Reporting Criteria

Vigilance System

An incident in EU vigilance systems is defined as "any malfunction or deterioration in the characteristics and/or performance of a device, as well as any inadequacy in the labeling or the instructions for use which, directly or indirectly, might lead to or might have led to the death of a patient, user or of other persons or has led to a serious deterioration in their state of health." [29]. In the EU, manufacturers must report medical device related serious adverse events and device failures, that might lead to or might have led to a death or

serious injury, to the competent authority (CA) in the nation of their occurrence. There is no legal requirement within the directives obliging users to have an active role in a vigilance system, but this area may be reinforced by separate advice from national regulatory bodies, as is the case of Finland, where National Supervisory Authority has issued national regulations on reporting serious adverse incidents for users and manufacturers. The duty to report applies to manufacturers and professional users of medical devices in Finland. Medical device serious adverse incidents must be reported to the authority within ten days of the user or manufacturer first becoming aware of the incident. The case of a near miss should be reported within thirty days [19, 21, 26, 29].

The core FDA requirement to manufacturers in the US vigilance systems requires reporting within 30 days of an awareness of a problem with a device. Key criteria for inclusion are devices that: (1) may have caused or contributed to a death or serious injury; or (2) have malfunctioned (and this device or a similar device that was marketed would be likely to cause or contribute to a death or serious injury, if the malfunction were to recur) [17]. The ONC requires that issues be taken up with vendors and developers and only if the vendor has not resolved the issue or there is dissatisfaction with solution should an ONC complaint be submitted [20].

Incident Reporting System

Patient safety reports usually describe the key data categories used to understand what happened, why it happened, and what were the consequences and reactions to the incident [27].

Analysis of the Incidents

Oversight System

This goal of the EU vigilance system is to be achieved by the evaluation of a reported incident, and where appropriate, dissemination of information could be used to prevent other occurrences of a similar event, or to alleviate the consequences of incidents. Suspected incidents are made known to the manufacturers and it is with their close involvement and co-operation that the implementation of the Field Safety Corrective Actions (FSCA) is made possible [19, 21, 26, 29].

In the US system there are no regulatory requirements to evaluate EHR system safety [17], and adverse outcomes associated with EHRs are not being systematically and consistently tracked. The regulatory data is stored in several databases supported by the FDA and ONC [19]. EHR certification alone does not guarantee that EHRs will be implemented and that they will work as planned [6, 14-18, 30]. If issues arise there is opportunity to: contact a vendor and this is followed by a formal complain with an ONC certification body [20, 31].

Incident Reporting System

Collected data is most commonly used for hazard identification and issuing of alerts, as well as for trends-cluster analysis. Risk, causal and systems analysis, are utilized in more mature US national reporting systems [27].

Summary

IRS are associated with voluntary reporting while oversight systems require that health professionals' are mandatory. Anonymity and confidentiality are important aspects of IRSs. Oversight systems do have usually more detailed HIT-specific

structures. Neither of these systems publically share data e.g., on a national level.

Discussion

This analysis provides a useful comparison between two types of reporting systems for technology-induced errors. The purpose and reporting criteria of IRS and ORS are not always clear for HIT users that are reporting these incidents. This paper may serve as guidance e.g., for clinicians by clarifying the major differences in these systems. In this context, our analyses reveal that there is a need for effective reporting about technology-induced errors. If there are multiple reporting systems with similar goals, health care professionals may become confused and this may negatively influence their willingness to report technology-induced errors. The issue of underreporting in both systems is a recognized phenomenon which requires further consideration. For example, underreporting exists in the US regulatory system. In the US there are only a small number of EHR incidents in differing databases so the total number of reports requires a review and analysis of multiple differing databases [7, 16–18]. EU studies are scarce. To address this underreporting of technology-induced errors there is a need to encourage clinician reporting, to develop reporting criteria for health professionals when reporting to IRS and ORS (in such a way as to avoid duplicate reports), and to outline the process and provide information about how issues are resolved for health professionals to improve trust in the safety of HIT.

Specifically, in the EU the details of the medical software directive are relatively unfamiliar for EHR users and leaders. The application of the requirements of the directive have not been clear. Not all stand alone medical software qualifies as a medical device [21, 24, 29]. This complexity of criteria highlights the fact that resources are needed to strengthen the clinicians' knowledge of reportable EHR-issues and consequently contributes to the reporting of incidents to improve the effectiveness of the regulatory system

When comparing oversight systems between the EU and US, there are challenges specifically with the US approach. Expert opinions of HIT oversight in the US and the HIT community should re-examine whether and how regulation of electronic health applications could foster patient safety [15]. Sittig and Singh proposed that in the US, there is a need to create a nationwide 'post-marketing' surveillance system to facilitate monitoring of HIT related safety events (i.e. technology-induced errors), and that methods and governance structures to support investigation of major HIT related safety events be developed [18]. Regarding both systems, there is an area still requiring consideration. Countries are at different stages of addressing technology-induced errors arising from HIT and there is considerable knowledge to be shared across the countries and internationally to improve HIT safety [32].

Conclusion

Criteria surrounding the types of reports that should be submitted to ORS and IRS systems are important so that health professionals know where to report such events and to avoid duplicate reporting (i.e., submitting one report across multiple databases). There is a need to provide information about HIT related safety events (and how they were resolved) to health professional users to encourage reporting of events and enhance their trust in the process of improving HIT safety.

References

- [1] Agboola, S., Bates, D. and Kvedar, J. (2016). Digital Health and Patient Safety. *JAMA*, 315(16), p.1697. doi: 10.1001/jama.2016.2402.
- [2] Denham, C.R., Classen, D.C., Swenson, S.J., Henderson, M.J., Zeltner, T. and Bates, D.W. (2013) 'Safe use of electronic health records and Health Information Technology systems', *Journal Patient Safety*, 9(4), pp. 177–189.
- [3] IOM (Institute of Medicine) (2012) 'Health IT and patient safety: building safer systems for safer care. Opportunities to improve safety in the use of health IT'. Washington, DC, USA: The National Academies Press; 2012.
- [4] Borycki, E. (2013) Technology-induced errors: where do they come from and what can we do about them? *Stud Health Technol Inform*, 194:20–6.
- [5] Magrabi, F., Aarts, J., Nohr, C., Baker, M., Harrison, S., Pelayo, S., Talmon, J., Sittig, D.F. and Coiera, E. (2013) 'A comparative review of patient safety initiatives for national health information technology', *International Journal of Medical Informatics*, 82(5), pp. e139–e148.
- [6] Magrabi, F., Ong, M.-S., Runciman, W. and Coiera, E. (2012) 'Using FDA reports to inform a classification for health information technology safety problems', *JAMIA*, 19(1), pp. 45–53. doi: 10.1136/amiainjnl-2011-000369.
- [7] Howell, A.-M., Burns, E.M., Hull, L., Mayer, E., Sevdalis, N. and Darzi, A. (2016) 'International recommendations for national patient safety incident reporting systems: An expert Delphi consensus-building process', *BMJ Quality & Safety*, pp. bmjqs-2015-004456.
- [8] Benn, J., Koutantji, M., Wallace, L., Spurgeon, P., Rejman, M., Healey, A. and Vincent, C. (2009). Feedback from incident reporting: information and action to improve patient safety. *Quality and Safety in Health Care*, 18(1), pp.11–21.
- [9] Brennan, T.A. (2000) 'The institute of medicine report on medical errors - could it do harm?', *New England Journal of Medicine*, 342(15), pp. 1123–1125.
- [10] Flink, E., Chevalier, C.L., Ruperto, A., Dameron, P., Heigel, F.J., Leslie, R., Mannion, J., Panzer, R.J. (2005) Lessons Learned from the Evolution of Mandatory Adverse Event Reporting Systems. In: Henriksen, K., Battles, J.B., Marks, E.S., Lewin, D.I., editors. *Advances in Patient Safety: From Research to Implementation (Volume 3: Implementation Issues)*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2005 Feb.
- [11] Marella, W.M., Sparnon, E. and Finley, E. (2014) 'Screening electronic health Record-Related patient safety reports using machine learning', *Journal of Patient Safety*, p. 1.
- [12] Warm, D. and Edwards, P. (2012) 'Classifying health information technology patient safety related incidents – an approach used in wales', *Applied Clinical Informatics*, 3(2), pp. 248–257.
- [13] Chai, K.E.K., Anthony, S., Coiera, E. and Magrabi, F. (2013) 'Using statistical text classification to identify Health Information Technology incidents', *JAMIA*, 20(5), pp.980–985.
- [14] Goodman, K.W., Berner, E.S., Dente, M.A., Kaplan, B., Koppel, R., Rucker, D., Sands, D.Z. and Winkelstein, P. (2011) 'Challenges in ethics, safety, best practices, and oversight regarding HIT vendors, their customers, and patients: A report of an AMIA special task force', *JAMIA*, 18(1), pp. 77–81.
- [15] Myers, R.B., Jones, S.L. and Sittig, D.F. (2011) 'Review of reported clinical information system adverse events in US food and drug administration databases', *Applied*

- Clinical Informatics, 2(1), pp. 63–74. doi: 10.4338/aci-2010-11-ra-0064.
- [16] De Moor, G., O'Brien J., Fridsma D., Bean, C., Devlies J, Cusack CM., Bloomrosen M., Lorenzi N, and Coorevits P.(2011) Policy brief on the current status of certification of electronic Health Records in the US and Europe. *Stud Health Technol Inform.* 2011;170:83-106.
 - [17] Singh, H., Classen, D.C. and Sittig, D.F. (2011) 'Creating an oversight infrastructure for electronic health Record–Related patient safety hazards', *Journal of Patient Safety*, 7(4), pp. 169–174. doi: 10.1097/pts.0b013e31823d8df0.
 - [18] Sittig, D.F., Classen, D.C. and Singh, H. (2014) 'Patient safety goals for the proposed federal health information technology safety center', JAMIA.
 - [19] EU (2007) Directive 2007/47/EC of the European Parliament and of the Council of 5 September 2007. Archived at: <http://www.webcitation.org/6lfm9YoLh>.
 - [20] Health IT.gov (2016) Archived at: <http://www.webcitation.org/6ms71VP9S>.
 - [21] EEC (1993) Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. Archived at: <http://www.webcitation.org/6lM7nIudx>.
 - [22] Arksey, H. and O'Malley, L. (2005) 'Scoping studies: Towards a methodological framework', *International Journal of Social Research Methodology*, 8(1), pp. 19–32.
 - [23] Furukawa, M.F., Eldridge, N., Wang, Y. and Metersky, M. (2016) 'Electronic health record adoption and rates of in-hospital adverse events', *Journal of Patient Safety*, p. 1.
 - [24] EU (2009) Council Recommendation on patient safety, including the prevention and control of healthcare associated infections, 2009/C 151/012. Available at <http://www.webcitation.org/6lp3DNpv4>.
 - [25] AMA (American Medical Association) (2009). The Physician's Guide to Patient Safety Organizations. American Medical Association; 2009. pp. 1–24. Available at <http://www.webcitation.org/6louKcZD4>.
 - [26] National Supervisory Authority for Welfare and Health In Finland. Available: <http://www.webcitaion.org/6lMAapqbN>
 - [27] Doupi, P. (2009) National Reporting Systems for Patient Safety Incidents. A review of the situation in Europe. National Institute for Health and Welfare (THL), Helsinki, Finland: Yliopistopaino, Report 13/2009, pp. 59.
 - [28] Haller, G., Myles, P.S., Stoelwinder, J., Langley, M., Anderson, H. and McNeil, J. (2007) Integrating incident reporting into an electronic patient record system. JAMIA, 14(2):175-81. Epub 2007 Jan 9.
 - [29] EC (2016) European Commission MEDDEV 2.1./6.Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory framework on medical devices. <http://www.webcitation.org/6lp3WD4UA>.
 - [30] Bowman, S. (2013) Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications. Perspectives in Health Information Management, 10(Fall):1c.
 - [31] HealthIT.gov Provider complaint process. <http://www.webcitation.org/6msFcY6XH>
 - [32] Kushniruk, A.W., Bates, D.W., Bainbridge, M., Househ, M.S. and Borycki, E.M. (2013) 'National efforts to improve health information system safety in Canada, the United States of America and England', *International Journal of Medical Informatics*, 82(5), pp. e149–e160. doi: 10.1016/j.ijmdinf.2012.12.006.

Corresponding author:

Sari Palojoiki
 University of Eastern Finland
 Preferred method of contacting: sari.palojoiki@hus.fi
 (Tel. +358504284179)

Factors Affecting the Time to Occurrence of Hospital-Acquired Pressure Ulcers Using EHR Data

Seul Ki Park^a, Hyeoun-Ae Park^{a, b}

^a College of Nursing, Seoul National University, Seoul, South Korea

^b Research Institute of Nursing Science, Seoul National University, Seoul, South Korea

Abstract

This study examined factors affecting the time to occurrence of hospital-acquired pressure ulcers using electronic health records (EHR) data in a critical care setting. The subjects were 202 patients who were admitted to a tertiary hospital between January 2015 and May 2016 and developed pressure ulcers. In total, 61 risk factors for pressure ulcers were extracted from nine different report forms and analyzed using univariate and multiple linear regression analyses. The univariate analysis identified 33 significant factors, and 11 factors remained significant in the multiple regression analysis. Patients with decreased consciousness and those on ventilators tended to develop pressure ulcers during the early stage of hospitalization. Patients taking sedative-hypnotics, anti-inflammatory drugs, steroids, and anticoagulants, those in renal failure, and those who consulted a nutritionist, or had a gastrointestinal or tracheostomy tube or diarrhea tended to develop pressure ulcers at a later stage of hospitalization.

Keywords:

Pressure Ulcer; Electronic Health Records; Risk Factors

Introduction

Hospital-acquired pressure ulcers (HAPUs) are a threat to patient health [1]. HAPUs may prolong the hospital stay and lead to increased medical costs [2, 3]. Therefore, it is important to prevent pressure ulcers. It is also important to treat patients with pressure ulcers properly to prevent complications, such as infection [4].

The incidence rates and time to the occurrence of HAPUs in a critical care setting vary across studies. The incidence of HAPUs ranges from 12.8 to 16.6% among patients in critical care settings [5-7], from 14.3 to 43.2% among intensive care unit (ICU) patients [8-10], and from 20.3 to 38.5% among trauma patients [11-13].

Regarding the time to HAPU occurrence, 59.1% of patients in a neurological ICU developed HAPUs within 3 days of their admission [14], and 28.8% of the patients admitted to a tertiary hospital developed them within 7 days of admission [15].

Many studies have examined the risk factors affecting pressure ulcer development. However, few studies have tried to identify risk factors related to the time to HAPU occurrence; one study explored the times to occurrence and closure and risk factors in patients with spinal cord injuries [16], and another study

explored risk factors for pressure ulcers among elderly patients early in the hospital stay [17].

With this background, we considered the factors affecting the time to HAPU occurrence among patients in a critical care setting with diverse diagnoses during their hospital stay. We also considered whether electronic health records (EHR) data could be used to answer this question.

Research Objectives

This study explored risk factors affecting the time to HAPU occurrence among adults admitted to a critical care setting.

Methods

Study design

This was a retrospective study that used EHR data to identify factors affecting the time to HAPU occurrence among patients admitted to an acute care hospital.

Setting and study subjects

The subjects were admitted to six wards in the neurology, neurosurgery, hemato-oncology, and oncology departments at a tertiary teaching hospital in Korea. The subjects were 202 adults who developed HAPUs from January 2015 to May 2016 in the study hospital. Patients with HAPUs were identified from pressure ulcer incident reports. For those patients who developed HAPUs more than once during their hospital stays, only the first HAPU was included in the study.

Data extraction from EHR

A literature review identified 61 pressure ulcer risk factors. To identify these risk factors in the EHR, nine documents were reviewed: the Braden Scale form, initial nursing assessment, clinical observation record, severity classification system, nursing activity sheet, nurses' notes, surgery record, doctors' order sheet, and laboratory test results. Risk factors were grouped based on how often each was measured and how variable the values were. For variables that were recorded only at admission (e.g., route of admission) or most frequently recorded at admission (e.g., body weight), data recorded at admission were extracted. For variables that were recorded after admission with values that do not vary by day or by shift (e.g., surgery or diagnosis), data recorded at the time when an event that occurred between admission and pressure ulcer occurrence were extracted. For variables that were recorded more than once and whose values varied by day or by shift, data recorded most recently before the pressure ulcer occurred were extracted. We

set the maximum data extraction allowance time by computing the average time difference between the last time the risk factor was recorded and the time of HAPU occurrence. For example, reports of pain appeared in the clinical observation records or nurses' notes. If a HAPU occurred, then the time difference between the most recent record of pain in the clinical observation record or nurses' notes and the time when the HAPU occurred was extracted, and the mean time difference was computed across patients. The mean time difference was 1 day. Thus, if pain was recorded within 1 day before the onset of a pressure ulcer, pain was considered to be present; otherwise, pain was considered to be absent.

If data were recorded in more than one form within the maximum data extraction time, data from the most recently completed form were selected. Data were extracted from the clinical data warehouse based on data-extraction guidelines prepared by the authors. Data quality was validated by reviewing the records of 10 patients selected randomly.

The recorded frequencies of the risk factors were counted, and factors with a relative frequency of less than 0.1% were not included in this study (e.g., diagnosis of narcotic intoxication, malnutrition, sepsis, and hypotension). Table 1 shows the data recording times and data sources by risk factors.

Analytical plan

The statistical analysis was performed with SPSS ver. 18.0 for Windows.

1. Descriptive analyses with the frequency, mean, and standard deviation were used to examine the characteristics of surgical and medical patients.
2. Univariate analyses were used to study factors affecting the time to HAPU occurrence.
3. Multiple stepwise regression analysis was performed to identify factors affecting the time to HAPU occurrence.

Results

Characteristics of the study subjects

The study included 202 patients with HAPUs: 128 surgical patients and 74 medical patients.

Table 2 presents the gender, age, diagnosis, Braden scale score on admission, time to HAPU occurrence from admission, place of occurrence, and body site for the surgical and medical patient groups.

Table 1 - Data recording time and data sources by risk factors

Data recording time	Risk Factors	EHR data source								
		1	2	3	4	5	6	7	8	9
On admission	Gender, age, smoking, route of admission, body weight	√								
	Braden scale score on admission		√							
Any time between admission and pressure ulcer occurrence	Surgery									√
	Diagnosis (cancer, fracture, diabetes mellitus, nervous system disease, urogenital disease, head injury, spinal cord injury, renal failure, cardiovascular disease, mental illness, respiratory disease)			√						
Until 1 day before pressure ulcer onset	Endotracheal tube, tracheostomy tube, oxygen therapy, restraint, gastrointestinal tube, motor disorder				√	√	√	√		
	Skin condition, immobilization devices							√		
	Hygiene management					√	√	√		
	Pain				√			√		
	Friction and shear		√							
	Albumin									√
Until 3 days before pressure ulcer onset	Nothing by mouth, edema, weakness							√		
	Cognitive impairment				√			√		
	Decreased consciousness				√		√	√		
	Foley catheter				√	√	√	√		
	Ventilator				√	√		√		
	Activity disorder						√	√		
Until 5 days before pressure ulcer onset	Malnutrition, skin moisture							√		
	Incontinence				√		√	√		
	Diarrhea				√			√		
	Nutrition consultation, medication (cardiac stimulants, sedative-hypnotics, analgesics, anti-inflammatory drugs, steroids, anxiolytics, or anticoagulants)			√						

EHR data sources: 1. Initial nursing assessment; 2. Braden Scale form; 3. doctors' order sheet; 4. clinical observation record; 5. nursing activity sheet; 6. severity classification system; 7. nurses' notes; 8. laboratory test result report; and 9. surgery record

Table 2 - Comparison of characteristics of surgical and medical patients

Characteristics	Surgical patients (n = 128)		Medical patients (n = 74)	
	n	%	n	%
Gender				
Male	76	59.4	43	58.1
Female	52	40.6	31	41.9
Age, mean \pm SD (years)	59.45 \pm 15.75		65.62 \pm 15.25	
Diagnosis				
Cancer	47	36.7	44	59.5
Nervous system disease	56	43.8	33	44.6
Cardiovascular disease	21	16.4	19	25.7
Diabetes mellitus	8	6.3	6	8.1
Renal failure	10	7.8	4	5.4
Spinal cord injury	8	6.3	2	2.7
Head injury	4	3.1	4	5.4
Mental illness	3	2.3	4	5.4
Fracture	5	3.9	2	2.7
Respiratory disease	3	2.3	2	2.7
Urogenital disease	1	0.8	1	1.4
Braden Scale score on admission				
Mild to very high risk	35	27.3	48	64.9
No risk	93	72.7	26	35.1
Time to HAPU occurrence (days)				
0-3	58	45.4	12	16.2
4-15	35	27.3	35	47.3
≥ 16	35	27.3	27	36.5
Place where pressure ulcer developed				
Operating room	73	57.0	0	0.0
Not the operating room	55	43.0	74	100.0
Body site of pressure ulcer				
Back of head	4	1.9	1	1.4
Cheek	11	5.1	0	0.0
Chest	28	13.0	0	0.0
Vertebra	14	6.5	2	2.8
Coccyx	42	19.4	49	69.0
Hip	20	9.3	18	25.4
Thigh	12	5.6	3	4.2
Lower leg	10	4.6	2	2.8
Ankle	14	6.5	0	0.0
Heel	11	5.1	0	0.0

Both groups had more males than females. The mean ages of the medical and surgical patients were 65.62 and 59.45 years, respectively. Of the surgical patients, 43.8% were diagnosed with nervous system diseases, and 59.5% of the medical patients were diagnosed with cancers. Most of the medical patients (64.9%) were classified as at risk on admission, whereas only 27.3% of the surgical patients were identified as at risk on admission. Surgical patients developed HAPUs during the early stage of the hospital stay, with 45.4% occurring within 3 days of admission. Medical patients developed HAPUs at a later stage of their hospital stays compared with surgical patients, with 47.3% developing at 4-15 days.

More than half of the surgical patients (57.0%) developed HAPUs in the operating room during surgery. The most common body site of the pressure ulcer for both surgical and medical patients was the coccyx. The second most common body site for pressure ulcers was the chest for surgical patients and the hip for medical patients.

Factors related to the time to HAPU occurrence

The univariate analyses identified 33 factors as significantly ($p < 0.05$) affecting the time from admission to HAPU occurrence (Table 3).

Cardiac stimulants, weight, Foley catheter, endotracheal tube, ventilator, oxygen therapy, nothing by mouth, restraint, decreased consciousness, cognitive impairment, hygiene management, pain, and surgery were negatively related to the length of time to HAPU occurrence from admission. This means that patients with these factors tended to develop HAPUs earlier in their hospital stays. Route of admission, Braden Scale score on admission, sedative-hypnotics, analgesics, anti-inflammatory drugs, steroids, anxiolytics, anticoagulants, renal failure, cancer, cardiovascular disease, incontinence, diarrhea, tracheostomy tube, gastrointestinal tube, weakness, nutrition consultation, malnutrition, activity disorder, and edema were positively related to the length of time to HAPU occurrence from admission. Patients with these factors tended to develop HAPUs later in their hospital stays.

Of the 33 factors included in the multiple stepwise regression analysis, 11 were significant, with combined explanatory power of 69.5% (Table 4). Patients with decreased consciousness and those on a ventilator developed HAPUs at an early stage of their hospital stays. Patients with sedative-hypnotics, anti-inflammatory drugs, steroids, anticoagulants, renal failure, nutrition consultation, diarrhea, gastrointestinal tube, and tracheostomy tube developed pressure ulcers at a later stage of their hospital stays.

Discussion

We divided our subjects into surgical and medical patient groups based on the fact that patients who are rendered immobile, such as those undergoing surgery, may be at greater risk. There were 128 surgical patients (63.4%). Of the 128 surgical patients, 73 (57.0%) developed pressure ulcers in the operating room. Surgery is an important risk factor for HAPU development in many studies [18-21]. This is likely because the patients are immobile during their surgery, and anesthesia prevents them from feeling the pain caused by compression [18]. In addition, blood loss, lowered body temperature, skin moisture due to irrigation, and warming therapy during

surgery are important factors in the development of pressure ulcers among surgery patients [21].

Table 3 - Factors significantly affecting the time to HAPU occurrence

Categories	Factors
Negatively related to the time to HAPU occurrence (13)	Ventilator, cognitive impairment, restraint, nothing by mouth, decreased consciousness, endotracheal tube, Foley catheter, pain, surgery, oxygen therapy, weight, hygiene management, cardiac stimulants
Positively related to the time to HAPU occurrence (20)	Nutrition consultation, sedative-hypnotics, analgesic, diarrhea, gastrointestinal tube, weakness, anxiolytics, anti-inflammatory, tracheostomy tube, anticoagulant, cancer, edema, route of admission, incontinence, cardiovascular disease, steroid, renal failure, admission Braden Scale score, malnutrition, activity disorder

The time to pressure ulcer occurrence after admission averaged 15.2 days. About one third of the pressure ulcers ($n = 70$, 34.7%) occurred within 3 days of admission. This is much lower than the value reported in a previous study, which found that 59.1% of neurological patients in ICU developed pressure ulcers within 3 days of admission [14]. This could be due to the fact that our subjects were more diverse and from many different wards.

Of the 70 patients who developed pressure ulcers within 3 days, 58 were surgical patients, and 12 were medical patients. The surgical patients were more likely to develop pressure ulcers during the initial stage of hospitalization compared with the medical patients. In our hospital, most surgical patients undergo surgery on the first or second day of admission, so surgical patients develop pressure ulcers in the early stage of hospitalization. In a previous study, most of the patients who underwent surgery were found to have developed pressure ulcers on the day of or within 2 days after surgery [20].

Among the surgical patients, 28 (13.0%) and 11 (5.1%) developed pressure ulcers on the chest or cheeks, respectively, whereas no medical patients had pressure ulcers on the chest or cheeks. Patients with a pressure ulcer on the chest, clavicle, or rib were those who had surgery in the prone position. Patients with pressure ulcers on the face were those who wore a medical device, such as an oxygen mask, during anesthesia [22].

The multiple regression analysis showed that decreased consciousness and being on a ventilator were the most significant factors affecting the development of pressure ulcers in the initial stage of admission. Decreased consciousness and being on a ventilator are directly related to general anesthesia. In another study, all of the surgical patients with pressure ulcers had general anesthesia [23]. In addition, the use of cardiac stimulants, such as vasopressors, during surgery was a significant factor affecting the development of pressure ulcers during the initial stage of admission. Postoperative pain and fasting for various tests or surgery were also related to the occurrence of pressure ulcers in the early stage of hospitalization.

Table 4 - Results of the multiple linear regression

Variable	β	T	p	Durbin-Watson	Adj. R ²
(constant)		14.050	<0.001		
Nutrition consultation	0.270	5.605	<0.001		
Ventilator	-0.309	-6.266	<0.001		
Sedative-hypnotics	0.183	3.934	<0.001		
Anti-inflammatory	0.175	4.211	<0.001		
Diarrhea	0.153	3.549	<0.001	1.905	0.695
Steroid	0.205	4.837	<0.001		
Renal failure	0.131	3.166	0.002		
Gastrointestinal tube	0.112	2.206	0.029		
Anticoagulant	0.107	2.484	0.014		
Decreased consciousness	-0.110	-2.267	0.025		
Tracheostomy tube	0.099	2.077	0.039		

Sedative-hypnotics, steroids, and anticoagulants were factors affecting the development of pressure ulcers at a later stage of hospitalization. Sedative-hypnotics include sleeping pills prescribed for medical patients, and dexamethasone-like steroids and anticoagulants for heparin therapy were used to treat neurological diseases.

A nutrition consultation also was an important factor in the development of pressure ulcers at a later time during hospitalization. Nutrition consultations are rarely requested for patients in the early stage of hospitalization, and are mainly requested for patients with nutritional deficiencies admitted to a long-term ward. Impaired nutritional intake was an important factor in the development of pressure ulcers in a study of medical patients [24].

A tracheostomy tube was a risk factor for pressure ulcer development at a later stage of hospitalization. Tracheostomy tubes are mainly used in patients with long-term hospitalization, whereas patients with surgical procedures have endotracheal tubes for ventilator applications, which was a risk factor affecting pressure ulcer development at an early stage of hospitalization.

In summary, most of the patients who developed pressure ulcers early in their hospitalization were patients who had general anesthesia during surgery. Patients who are admitted for surgery have a high risk of developing pressure ulcers due to surgery, although they were assessed as low risk for pressure ulcers on admission.

In this study, we did not use the incident report as a data source to identify risk factors for pressure ulcers, but the incident reports can be connected with EHR to deepen the analysis of risk factors.

Conclusions

This study examined the factors affecting the time to HAPU occurrence during the hospital stay. These findings can be

used to help nurses to develop nursing interventions to recognize and prevent HAPUs.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2015R1A2A2A01008207).

References

- [1] Stifter, J., et al. Nurse continuity and hospital-acquired pressure ulcers: a comparative analysis using an electronic health record" big data" set. *Nursing Research* 64.5 (2014), 361–371.
- [2] Cho, K.H., Jun, K.J., Bok, S.K., Hong, J.H., Lee, H., Park, N.K., Choe, H.S. Epidemiologic study on clinical features of patient with pressure ulcer—A prospective study. *Journal of Korean Academy of Rehabilitation Medicine* 29.1 (2005), 122–127.
- [3] Russo, C. Allison, Claudia Steiner, and William Spector. Hospitalizations related to pressure ulcers among adults 18 years older, 2006. Statistical brief# 64. *Agency for Healthcare and Research Quality* 2008 (2012).
- [4] Reddy, M., Gill, S. S., Rocho, P. A. Preventing pressure ulcers: a systematic review. *JAMA* 296.8 (2006), 974–984.
- [5] Bergstrom, N., Braden, B., Kemp, M., Champagne, M., Ruby, E. Multisite study of incidence of pressure ulcers and the relationship between risk level, demographic characteristics, diagnoses, and prescription of preventive interventions. *Journal of the American Geriatrics Society* 41.1 (1996), 22–30.
- [6] Halfens, R.J.G., Van Achterberg, T., Bal, R.M. Validity and reliability of the Braden scale and the influence of other risk factors: a multicentre prospective study. *International Journal of Nursing Studies* 37.4 (2000), 313–319.
- [7] Pancorbo Hidalgo, P. L., Garcia Fernandez, F. P. Risk factors for the development of pressure ulcers among hospitalized elderly patients. *Gerokomos* 12.4 (2001), 175–184.
- [8] Sayar, S., Turgut, S., Dogan, H., Ekici, A., Yurtsever, S., Demirkan, F., Doruk, N., Tasdelen, B. Incidence of pressure ulcers in intensive care unit patients at risk according to the Waterlow scale and factors influencing the development of pressure ulcers. *Journal of Clinical Nursing* 18.5 (2009), 765–774.
- [9] Suriadi, Sanada, H., Sugama, J., Kitagawa, A., Thigpen, B., Kinoshita, S., Murayama, S. Risk factors in the development of pressure ulcers in an intensive care unit in Pontianak, Indonesia. *International Wound Journal* 4.3 (2007), 208–215.
- [10] Rose, P., Cohen, R., Amsel, R. Development of a scale to measure the risk of skin breakdown in critically ill patients. *American Journal of Critical Care* 15.3 (2006), 337–341.
- [11] Watts, D., Abrahams, E., MacMillan, C., Sanat, J., Silver, R., VanGorder, S., Waller, M., York, D. Insult after injury: pressure ulcers in trauma patients. *Orthopaedic Nursing* 17.4 (1998), 84–91.
- [12] Rademakers, L.M.F., Vainas, T., van Zutphen, S. W. A. M., Brink, P. R. G., van Helden, S. H. Pressure ulcers and prolonged hospital stay in hip fracture patients affected by time-to-surgery. *European Journal of Trauma and Emergency Surgery* 33.3 (2007), 238–244.
- [13] Salzberg, C. A., Byrne, D. W., Kabir, R., van Nieuwerburgh, P., Cayten, C. G. Predicting pressure ulcers during initial hospitalization for acute spinal cord injury. *Wounds. A Compendium of Clinical Research and Practice* 11.2 (1999), 45–57.
- [14] Im M.J., Park H.S. A study on the pressure ulcers in neurological patients in intensive care units. *Journal of the Korean Academy of Fundamentals of Nursing*, 13.2 (2006), 190–199.
- [15] Kim, K. N., Kang, K. J., Lee, H. S., Shin, Y. H., Kim, S. K., & Park, K. H. Survey on pressure ulcers and influencing factors of stage change in acute university hospital inpatients. *J Korean Clin Nurs Res*, 17.3 (2011), 433–442.
- [16] H van der Wielen, H., et al. Hospital-acquired pressure ulcers in spinal cord injured patients: time to occur, time until closure and risk factors. *Spinal Cord* 54.9 (2016), 726–731.
- [17] Baumgarten, M., et al. Pressure ulcers among elderly patients early in the hospital stay. *The Journals of Gerontology Series A. Biological Sciences and Medical Sciences* 61.7 (2006), 749–754.
- [18] Lindgren, M., et al. Immobility—a major risk factor for development of pressure ulcers among adult hospitalized patients: a prospective study. *Scandinavian Journal of Caring Sciences* 18.1 (2004), 57–64.
- [19] Chen, H.L., Chen X.Y., Wu J. The incidence of pressure ulcers in surgical patients of the last 5 years: a systematic review. *Wounds* 24.9 (2012), 234–241.
- [20] Aronovitch, S. A. Intraoperatively acquired pressure ulcer prevalence: a national study. *Journal of Wound Ostomy & Continence Nursing* 26.3 (1999), 130–136.

- [21] Lumbley, J. L., Ali, S. A., Tchokouani, L. S. Retrospective review of predisposing factors for intraoperative intra operative pressure ulcer development. *Journal of Clinical Anesthesia* 26.5 (2014), 368–374.
- [22] Black, J. M., et al. Medical device related pressure ulcers in hospitalized patients. *International Wound Journal* 7.5 (2010), 358–365.
- [23] Park, H.S., Park, K.Y., Yu, S.M. Factors influencing the development of pressure ulcers in surgical patients. *Journal of Korean Academy of Nursing* 35.1 (2005), 125–134.
- [24] Berlowitz, D. R., Wilking, S. V. Risk factors for pressure sores. A comparison of cross-sectional and cohort-derived data. *Journal of the American Geriatrics Society* 37.11 (1989), 1043–1050.

Address for correspondence

Hyeoun-Ae Park, PhD

College of Nursing, Seoul National University, 103 Daehak-ro, Jongno-gu, Seoul 110-799, South Korea. Tel.: +82-2-7408827, Fax: +82-2-765-4103, E-mail: hapark@snu.ac.kr

Precision of EMR Data: The Case for a Drug and Alcohol Service

Siyu Qian^a, Esther Munyisia^b, Ping Yu^a, David Hailey^a, David Reid^c

^a School of Computing and Information Technology, University of Wollongong, Wollongong, New South Wales, Australia

^b School of Psychology and Public Health, La Trobe University, Melbourne, Victoria, Australia

^c Illawarra Shoalhaven Local Health District, Drug and Alcohol Service, New South Wales, Australia

Abstract

Error-laden data can negatively affect clinical and operational decision making, research findings and funding allocation. This study examined the number and types of data errors in an electronic medical record (EMR) system in a Drug and Alcohol service. Specifically, errors in service data were examined. Three months after the implementation of the EMR system, 9,379 errors were identified from ten error reports generated between March 2015 and May 2016. The errors were grouped into four types: mismatched data fields (60.5%), duplicate medical record error (3.2%), date/time error (8.8%) and blank field error (27.4%). The errors can be prevented by adding functions, such as alert messages in the EMR system. How and why the errors occur need to be investigated in future studies.

Keywords: Electronic Health Records; Informatics; Medical Order Entry Systems

Introduction

The adoption and use of health information systems (HIS) has made large amounts of digital data available for use in clinical and operational decision making [1], research [2] and funding allocation [3]. However, data errors may jeopardize the realization of these purposes. For example, Ward et al. found that data errors in HIS time stamps can compromise the ability of an emergency department to accurately determine its operational performance [1].

Inherent data errors in clinical research databases may negatively impact the research findings [4, 5]. In a study comparing the estimation results of mortality rates using an error-free database and an error-seeded database, authors found that the overestimated mortality rates are typical results of using the latter database [6]. The estimates can be more than double the true value [6].

Error-laden data can also lead to misallocation of healthcare funding. In a study examining the consequences of miscoding in a hospital in Australia, authors found that about 16% of inpatient cases discharged from a specialized surgical unit during a six-month time period were miscoded. This led to an approximately \$575,000 underpayment to the hospital [3].

Understanding the types and number of errors in an organization's HIS is thus useful for managers to develop strategies to prevent errors. This will ensure that the HIS truly supports organizational performance measurements, decision making, research and funding allocation.

Data errors identified by previous studies include time stamp error [1], miscoding [3], missing data [7], data transfer error [8], spelling error [9], duplicate records [10, 11], drop-down menu selection error [10] and inconsistencies between data fields [12].

Causes of these errors are related to HIS design and how HIS are used [8]. For example, the drop-down menu selection error may be caused by too many items in the drop-down list or items being too close together [13]. Spelling errors may be attributed to healthcare providers documenting in a rush without proofreading [9].

Alcohol is one of the major risks for both physical and social health. Excessive consumption of alcohol can cause a wide range of harm including road accidents, domestic and public violence, family breakdown, crime, liver disease and brain damage [14]. Between 2014 and 2015, more than 115,000 Australians received over 170,000 treatment episodes from publicly-funded Drug and Alcohol (D&A) service [15].

In 2013, the public health in New South Wales (NSW), Australia moved from lump-sum funding of D&A services to funding based on treatment activities. The effectiveness of this new model of funding is substantially relied on the precision of data recorded in a HIS. Inaccurate recording may not only affect the amount of funding allocated to a D&A service, but also the managerial decisions made using this data, for example, in policy making, service planning, research and education [16]. To our knowledge, however, no study has investigated the precision of D&A service data under this model of funding.

Therefore, this study aimed to investigate the number and types of errors in service data in a D&A service in NSW, Australia.

Methods

Study setting

This study was conducted in a D&A service in NSW, Australia.

Data source

An electronic medical record (EMR) system was implemented in December 2014. Service activity data was extracted by performance unit staff from the EMR system and uploaded to Web Non-Admitted Patient (WebNAP), a system that reports outpatient activities to NSW Ministry of Health for use in state health policy decision making including funding.

The WebNAP system matched the uploaded activity data with its predefined activity classification. When the data did not match the classification, the WebNAP system identified an error and automatically recorded it in an error report. For example, the WebNAP activity classification showed that D&A was a community service, therefore D&A healthcare providers should always choose community as their setting type. If a provider chose hospital, an error occurred.

Datasets

Ten error reports generated between March 2015 and May 2016 were provided by the information manager in the D&A service, three to fourteen months after implementation of the system. This relates to approximately 150,000 records per year for 60,000 patients. On past experiences, errors would be expected in 7% of the records with more than one error in 2% of these records.

An error report contains the following data fields in an Excel spreadsheet: clinic name, healthcare worker's name, client name, client medical record number, error description, appointment date/time, service date/time, referral date/time, referral receipt date/time, service type, provider type, setting type, modality of care, financial class, Department of Veterans' Affairs (DVA) card type and DVA card number.

Names of healthcare workers and clients were deidentified to maintain confidentiality.

Data analysis

Data on error description was extracted for analysis by the researchers. The errors are labeled based on the feedback from the WebNAP system and were further grouped to higher level categories. The number of each type of error was counted.

Ethics approval

Ethics approval to conduct this study was granted by the joint Health and Medical Human Research Ethics Committee of the University of Wollongong and the Illawarra & Shoalhaven Local Health District. Access to the error reports was granted by the service manager of the D&A service. Consent was obtained from healthcare providers for their error reports to be used in this study.

Results

Types of Errors in D&A Service Data

The identified errors were grouped into four types: 'mismatched data fields', 'duplicate record', 'date/time error' and 'blank field' (see Table 1).

There were two types of 'mismatched data fields'. One was a 'service option error' and the other was a 'DVA information error.' 'Service option error' occurred when one of the three data entry fields, provider type, modality of care and setting type did not match what was set up in the WebNAP classification. Provider type described the job role of a healthcare provider e.g., a registered nurse, a psychologist or a counselor. Modality of care was the means for delivery of a service; e.g., telephone, email or face-to-face meeting between a healthcare provider or a group of healthcare providers and a client. Setting type was the location where the service was provided; e.g., hospital, community or home.

The 'DVA information error' occurred when the information provided in the two data entry fields, DVA card details and financial classification, did not match with each other. There were two situations. One was a 'DVA card details supplied but financial classification is not DVA' and the other was 'financial classification is DVA but missing DVA card details'.

'Duplicate medical record' occurred when more than one encounter was created for a client in the EMR system. However, a client could only have one active encounter at a time.

The 'date/time error' was related to referral date/time, referral receipt date/time, service start date/time and service end date/time. These date/times include time information. The

logical order of these date/times must be as follows: the referral date/time should be earlier than the referral receipt date/time which should be earlier than the service start date/time and the service end date/time. When this order was turned around, an error occurred.

The last data error type was 'blank field' which occurred when a data entry field was blank. The identified blank fields were provider type, funding source, financial classification and address fields including post code, suburb and street.

Table 1– Number and types of errors in service data in the EMR system in a D&A service.

Error types	Number	%
Total	9379	100
Mismatched data fields	5675	60.5
Service option error (provider type, modality of care or setting type does not match what is in WebNAP classification.)	5650	60.2
*DVA information error (DVA card details supplied but financial classification is not DVA, or financial classification is DVA but missing DVA card details)	25	0.3
Duplicate medical record	303	3.2
A second encounter is created in the EMR, but a client can only have one active encounter at a time		
Date/time error	834	8.8
Service start date/time is after service end date/time	591	6.3
Referral receipt date/time is after service date/time or before referral date/time	145	1.5
Referral date/time is after service start date/time	98	1
Blank field	2567	27.4
Blank provider type	2473	26.4
Blank post code	36	0.4
Blank suburb	21	0.2
Blank street	21	0.2
Blank funding source code	10	0.1
Blank financial classification	6	0.1

*DVA: Department of Veterans' Affairs

Number of Errors in Each Error Type

Overall, 9,379 errors were identified from the error reports (see Table 1). A total of 60.5% of the errors were 'mismatched data fields', of which, the majority was a 'service option error.' The error that occurred most frequently accounted for 27.4% and were 'blank field' with 'blank provider type.'. 'Date/time error' accounted for 8.8% of the total number of errors. 'Service start date/time is after service end date/time' was the major error which accounted for 6.3%. 'Duplicate medical record' was the least frequently occurring error which accounts for 3.2% of the total errors.

Discussion

As errors in operational data can be pervasive in the immediate period after the implementation of HIS [7], this study focused on analyzing error reports for an EMR system three to 14 months after its implementation. The large amount of errors found in this study may be due to the learning curve of healthcare providers. Further study will investigate trend in data errors over time.

We classified errors into four types. This is useful for investigating causes of error and developing different prevention and mitigation strategies required for different error types.

Causes of these errors might be related to the EMR system design issue, how the system was used, the environment in which it was used or a combination of them. For example, this study found that ‘service option error’ was the most common error in the D&A service data and wrong selection of ‘setting type’ was one immediate reason for it. In the WebNAP reporting system, the D&A service was mapped as a community setting. That means all healthcare providers working in the D&A service must select ‘community’ for setting type, regardless where the service was provided. This would ensure that funding was allocated to D&A service.

Some of the healthcare providers, although employed to work for the D&A service, were responsible for providing this specialist service in a hospital setting. A D&A healthcare provider might select ‘hospital’ for the setting type, because he or she provided the service there. This would result in a ‘service option error’ because the setting type of ‘hospital’ was not mapped in the WebNAP system for the D&A service. If this error was not corrected, funding would be misplaced to the hospital setting, instead of D&A service.

The term ‘setting type’ may confuse healthcare providers, especially new employees, on whether it means the location of service provided or the healthcare service by which the healthcare provider was employed. This could be one reason for the error to occur.

Another reason might be the design issue of the WebNAP system. The system used ‘setting type’ to determine which healthcare service would get funding. Actually, it could have used the information about the healthcare service by which the healthcare provider was employed to determine the correct ‘setting type’. This suggested that the system designer did not really understand how healthcare providers worked and how funding was allocated. Extraction of correct data elements would eliminate the error and ensure correct funding allocation.

A mix of system design issues and the environmental condition under which end users use the system may contribute to the occurrence of error. This study found that ‘date/time error’ accounted for almost 9% of the total number of errors. Date/time data were required from three forms that D&A healthcare providers used in the EMR system: intake form, assessment form and clinical note. The intake form was used when a client first contacted the D&A service. The assessment form was used to assess the person after intake. The clinical note was used in the subsequent visits of the client. The entire journey of the person with the D&A service from intake to discharge is an ‘encounter’. Each contact of the person with the D&A service may contain several ‘services’, e.g. counselling, rehabilitation or supervised medication administration. The three forms were used at different stages of the encounter. The intake form and the assessment form were used in the beginning of this encounter. The clinical note was used multiple times until the closure of the encounter.

Within each form, the service start date/time must be before the service end date/time. However, the way to record service start and end date/time is different among the three forms (see Table 2). The intake form automatically populated service start date/time from the computer, but required healthcare providers to manually enter end date/time. The assessment form required both start and end date/time to be manually entered. The clinical note required a manual entry of start date/time, but automatically filled in the end date/time. This mixed ways of recording date/time opens opportunities for error.

Another reason for the error to occur might be the environmental condition under which healthcare providers recorded data. Sitting with a client and recording data in a computer at the same time may increase the probability of making error. Automatic time recording may have better data accuracy than manual time recording [17]. A check of date/time by the EMR system at the data entry stage may also help to prevent the erroneous data from being recorded.

Table 2– How service start and end date/time is recorded.

Forms	Service start date/time	Service end date/time
Intake form	Auto-populated, can be adjusted manually	Manually entered
Assessment form	Manually entered	Manually entered
Clinical note	Manually entered	Auto-populated, can be adjusted manually

A lack of alert message function in the system may open opportunity for error. For example, although the percentage of ‘duplicate medical record’ is not high in this study, duplicate records can mislead healthcare providers in clinical decision making because they may miss important information that exists in a different record [11]. This error may also cause confusion in information retrieval [18]. Disruptive pop-up alert message [4] built into an EMR system may help to decrease this error.

The ‘blank provider type’ is the second common error found in this study. This error might be caused by system dysfunction. It may also be due to the inability of the system administrator to keep up with the workload of correctly mapping the providers in the WebNAP system because of the high turnover of D&A staff, or lack of information on changes in roles for staff members or new staff in the service. Further study is needed to investigate how and why this error occurred.

Since the service data was first extracted by the performance unit staff before submitting to WebNAP, it was possible that the errors reported in this study could also be injected during the process of data extraction.

With the increased use of EMR, research using EMR data has been prospering, for example in EMR phenotyping [19], clinical workflow modeling [20] and disease prediction [21]. The strength of evidence from the secondary analysis of EMR data can be hindered by errors contained in these data [4, 5]. In the case mentioned in the introduction section, the study comparing the estimation results of mortality rates using an error-free database and an error-seeded database, the analysis result using the error-laden data was more than double the true results using the error-free data [6]. This shows the negative impact of erroneous data in reducing accuracy of data

analytics and lead to invalid findings. Therefore, data errors and their causes need to be identified, reduced or eliminated to ensure high data quality to provide accurate evidence for research and health decision making [22].

Limitation

We did not analyze all the error reports generated after the implementation of the EMR system since there was no health information manager at the D&A service for a period of three months. Consequently, no error reports were downloaded in this period. However, the errors made during that time period rolled over to the following months' reports, so we think that we had all errors analyzed. Directly interviewing the healthcare providers who made the errors would give further insights about the reasons for the errors made, which would be beneficial for learning to prevent errors in the EMR system.

Conclusion

Error-laden data can jeopardize clinical and operational decision making [1], research findings [6] and funding allocation [3]. To manage errors, it is paramount for D&A service to understand the nature and extent of error and the environment that induce error [23].

This study investigated the number and types of errors in service data in a D&A service. The identified errors were grouped into four error types: 'mismatched data fields' (60.5%), 'duplicate medical record' (3.2%), 'date/time error' (8.8%) and 'blank field' (27.4%). The top three most frequent sub errors were 'service option error,' 'blank provider type' and 'service start date/time is after service end date/time.' The results from this study underscore the importance of understanding errors in EMR data. Further study will investigate the trend in data errors overtime and how and why errors occur in the EMR system.

Acknowledgements

We are grateful for the healthcare providers at the D&A service who provided consent for their error reports to be used in this study. We thank the director and service managers for their support and approval of this study. The health information manager Mr. Will Fernandez is acknowledged for supporting the data collection process.

References

- [1] M.J. Ward, W.H. Self, and C.M. Froehle, Effects of Common Data Errors in Electronic Health Records on Emergency Department Operational Performance Metrics: A Monte Carlo Simulation, *Academic Emergency Medicine* **22** (2015), 1085-1092.
- [2] M.A. Wildeman, J. Zandbergen, A. Vincent, C. Herdini, J.M. Middeldorp, R. Fles, O. Dalesio, E. Van der Donk, and I.B. Tan, Can an online clinical data management service help in improving data collection and data quality in a developing country setting?, *Trials* **12** (2011).
- [3] P. Cheng, A. Gilchrist, K.M. Robinson, and L. Paul, The risk and consequences of clinical miscoding due to inadequate medical documentation: a case study of the impact on health services funding, *Health Information Management Journal* **38** (2009), 35-46.
- [4] K. Haerian, J. McKeeby, G. Dipatrizio, and J.J. Cimino, Use of clinical alerting to improve the collection of clinical research data, *AMIA. Annual Symposium proceedings / AMIA Symposium*. *AMIA Symposium* 2009) 218-222.
- [5] S. Goldberg, A. Niemierko, and A. Turchin, Analysis of data errors in clinical research databases, in: *AMIA, Citeseer*, 2008.
- [6] S. Gallivan, J. Stark, C. Pagel, G. Williams, and W.G. Williams, Dead reckoning: can we trust estimates of mortality rates in clinical databases?, *European Journal of Cardio-Thoracic Surgery* **33** (2008), 334-340.
- [7] M.J. Ward, C.M. Froehle, K.W. Hart, and C.J. Lindsell, Operational data integrity during electronic health record implementation in the ED, *American Journal of Emergency Medicine* **31** (2013), 1029-1033.
- [8] D.F. Sittig and H. Singh, A red - flag - based approach to risk management of EHR - related safety concerns, *Journal of Healthcare Risk Management* **33** (2013), 21-26.
- [9] F. Hussain and U. Qamar, Identification and correction of misspelled drugs' names in electronic medical records (EMR), in: *ICEIS 2016 - Proceedings of the 18th International Conference on Enterprise Information Systems*, 2016, pp. 333-338.
- [10] K.E. Walsh, W.G. Adams, H. Bauchner, R.J. Vinci, J.B. Chessare, M.R. Cooper, P.M. Hebert, E.G. Schainker, and C.P. Landrigan, Medication errors related to computerized order entry for children, *Pediatrics* **118** (2006), 1872-1879.
- [11] A.B. McCoy, A. Wright, M.G. Kahn, J.S. Shapiro, E.V. Bernstam, and D.F. Sittig, Matching identifiers in electronic health records: implications for duplicate records and patient safety, *BMJ quality & safety* **22** (2013), 219-224.
- [12] H. Singh, S. Mani, D. Espadas, N. Petersen, V. Franklin, and L.A. Petersen, Prescription errors and outcomes related to inconsistent information transmitted through computerized order entry: a prospective study, *Archives of Internal Medicine* **169** (2009), 982-989.
- [13] R. Khajouei and M. Jaspers, The impact of CPOE medication systems' design aspects on usability, workflow and medication orders, *Methods of Information in Medicine* **49** (2010), 3.
- [14] Australian Institute of Health and Welfare, Alcohol and other drugs, 2016. Available: <http://www.aihw.gov.au/alcohol-and-other-drugs/>
- [15] Australian Institute of Health and Welfare, Alcohol and other drug treatment services (AODTS), 2016. Available: <http://www.aihw.gov.au/alcohol-and-other-drugs/aodts/>
- [16] B.G. Audit Commission, Data quality matters: learning from audit commission work on data quality in local public services, in, 2015.
- [17] R.A. Marjamaa, P.M. Torkki, M.I. Torkki, and O.A. Kirvelä, Time accuracy of a radio frequency identification patient tracking system for recording operating room timestamps, *Anesthesia and Analgesia* **102** (2006), 1183-1186.
- [18] S. Qian and P. Yu, Fitting clinical workflow: The case for wound care in a residential aged care home, in: *Studies in Health Technology and Informatics*, H. Grain, F. Martin-Sanchez, and L.K. Schaper, eds., IOS press, Melbourne, Australia, 2014, pp. 130-136.
- [19] T. Zheng, W. Xie, L. Xu, X. He, Y. Zhang, M. You, G. Yang, and Y. Chen, A machine learning-based framework to identify type 2 diabetes through electronic health records, *International Journal of Medical Informatics* **97** (2017), 120-127.
- [20] Y. Chen, W. Xie, C.A. Gunter, D. Liebovitz, S. Mehrotra, H. Zhang, and B. Malin, Inferring Clinical Workflow

Efficiency via Electronic Medical Record Utilization, *AMIA Annual Symposium Proceedings* 2015, 416-425.

- [21] R. Kop, M. Hoogendoorn, L.M. Moons, M.E. Numans, and A. ten Teije, On the advantage of using dedicated data mining techniques to predict colorectal cancer, in: *Conference on Artificial Intelligence in Medicine in Europe*, Springer, 2015, pp. 133-142.
- [22] S. Hoffman and A. Podgurski, Big Bad Data: Law, Public Health, and Biomedical Databases, *Journal of Law, Medicine and Ethics* **41** (2013), 56-60.
- [23] R.L. Helmreich, On error management: lessons from aviation, *British Medical Journal* **320** (2000), 781.

Address for correspondence

Associate Professor Ping Yu
School of Computing and Information Technology, University of
Wollongong, Wollongong, New South Wales, 2522, Australia.
Work phone: +61 02 4221 416
Email: ping@uow.edu.au

Adverse Drug Event Monitoring with Clinical and Laboratory Data Using Arden Syntax

Andrea Rappelsberger^a, Klaus-Peter Adlassnig^{a,b}, Jeroen S. de Bruin^{a,b},
Manuela Plössnig^c, Jochen Schuler^d, Christina Hofer-Dückelmann^e

^a Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

^b Medexter Healthcare GmbH, Borschkegasse 7/5, A-1090 Vienna, Austria

^c Salzburg Research Forschungsgesellschaft mbH, Jakob Haringer Strasse 5/3, A-5020 Salzburg, Austria

^d Paracelsus Medical Private University Salzburg, Institute of General Practice, Family Medicine and Preventive Medicine, Strubergasse 21, A-5020 Salzburg, Austria

^e Landesapotheker am St. Johanns-Spital Salzburg, Müllner Hauptstrasse 50, A-5020 Salzburg, Austria

Abstract

In times of steadily increasing numbers of administered drugs, the detection of adverse drug events (ADEs) is an important aspect of improving patient safety. At present only about 1–13% of detected ADEs are reported. Raising the number of reported ADEs will result in greater and more efficient support of pharmacovigilance. Potential ADEs must be identified early. In the iMedication system, which is a rule-based application, triggers are used for computerized detection of possible ADEs. Creating a pilot system, we defined the relevant use cases hyperkalemia, hyponatremia, renal failure, and over-anticoagulation; knowledge bases were implemented in Arden Syntax for each use case. The objective of these knowledge bases is to interpret patient-specific clinical data and generate notifications based on a calculated ADE risk score, which may indicate possible ADEs. This will permit appropriate monitoring of potential ADE situations over time in the interest of patient care, quality assurance, and pharmacovigilance.

Keywords:

Drug-Related Side Effects and Adverse Reactions; Drug Monitoring; Expert Systems.

Introduction

Medical errors or unintentional acts of omission or commission, or those that do not achieve their intended outcome [1] – such as inappropriate medication – do occur in hospitals. A recent study stated that medical errors are the third leading cause of death in the U.S. [2], making it more important than ever to prevent and mitigate medical errors, especially those causing damage to patients [3].

The fact that drugs are being administered in increasing numbers signifies a greater potential for drug-related harm, including adverse drug events (ADEs). Traditionally ADEs are tracked and reported on a voluntary basis. Hence the success of ambulatory error reporting systems has been limited; approximately 10–20% of medication errors and only 1–13% of detected ADEs are reported [4]. Additionally, the process of ADE detection consumes considerable resources in terms of time and money. Studies have shown that as many as 6% of all hospital admissions are due to ADEs, and this

number is three- fold higher among elderly patients [5, 6]. Moreover, about 50% of these prescribing errors and ADEs are deemed avoidable [7].

Hospitals need a more efficient mode of quantifying the degree and severity of ADEs, such as automated or computerized detection. Identification of severe ADEs as well as measuring their frequency will enable pharmacists and physicians to take corrective measures.

iMedication supports the process of pharmacovigilance – the pharmacological science relating to the detection, assessment, understanding and prevention of adverse effects, especially the long-term and short-term side effects of medication [8] – by identifying potential ADEs. Using automated tools such as iMedication, it is possible to reduce the number and severity of ADEs over time, identifying potential ADEs as early as possible, supporting plausibility checks on suspected ADEs, and reporting verified ADEs in an appropriate and standardized manner. Furthermore, it can inspire physicians and pharmacists to report ADEs and – last but not least – save time and money during the reporting process.

Existing approaches for computerized ADE detection employ methods such as data mining [9] and decision trees [10] to automatically generate ADE detection rules. Another strategy is to utilize the rich semantics of ontologies such as SNOMED CT [11, 12], and apply it to the detection of ADEs via semantic querying and reasoning. Others approach the task by the automatic creation of rules with the aid of product label parsing [13]. In the iMedication project, we integrate the operative knowledge of local and remote experts by linking distributed knowledge repositories and manually derive specific rules from expert knowledge. This enables us to specify complex rules for the identification of ADEs. The system reports detected ADEs according to their severity. The reports additionally include an explanation as to how the knowledge base came to its conclusion to report an ADE. Furthermore, the report provides information that helps the physician or clinical pharmacist to take corrective therapeutic measures. If necessary, a report is forwarded to the Austrian Agency for Health and Food Safety (AGES), the agency responsible for pharmacovigilance in Austria. The workflow of ADE identification, verification, and reporting is shown in Figure 1.

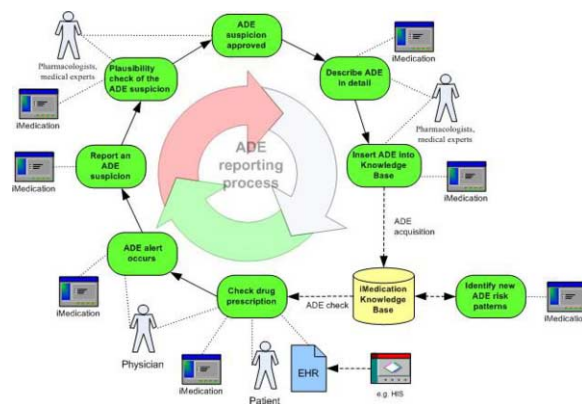


Figure 1– Workflow of adverse drug event (ADE) detection and reporting.

In the present paper, we report the results of a pilot study on effectiveness performed in 2012. Using data on patients admitted to the University Hospital of Salzburg (UHS) in 2007 and 2011, we analyze the sensitivity and specificity of the system in detail.

Methods

Theoretical foundations

The iMedication system is founded on the principles of the “Institute for Healthcare Improvement (IHI) Global Trigger Tool” method [3] and Morimoto’s classification [14] for the detection of possible ADEs.

The IHI Global Trigger Tool for Measuring Adverse Events is a method of identifying adverse events – especially those causing harm – and measuring the rate of adverse events over time. The method employs triggers – clues on possible adverse events – to track adverse events, including ADEs. However, the tool is not meant to identify all adverse events, but rather performs a retrospective review of a random sample of inpatient data [3].

According to Morimoto et al. [14], irregular use of medication (referred to as incidents) can be classified in many ways: actual ADEs vs. potential; preventable vs. non-preventable; ameliorable vs. non-ameliorable; and errors vs. non-error. According to this method, an ADE is regarded as an injury due to medication.

In general, incidents are identified by collecting practice data, soliciting incident reports from patient caregivers, and surveying patients directly. These data are then independently reviewed by patient caregivers using various triggers, such as:

- Symptoms or actions that suggest a (potential) ADE or medication error, such as a new rash or new diarrhea.
- Diagnoses associated with (potential) ADEs or medication errors, such as poisoning by drugs.
- The use of specific drugs that suggest an ADE may have occurred.
- Drug combinations are known to cause ADEs or the use of duplicate drugs.

- Combinations of drugs and symptoms that might indicate a (harmful) reaction to the drug, such as diarrhea or eruption due to antibiotics.
- Combinations of drugs and patient diagnoses, such as bleeding and antiplatelet agents or warfarin.
- Combinations of drugs and other factors such as patient age or sex, or pregnancy.
- Laboratory triggers, such as microbiology results that show an improper use of antibiotics.

Study design, setting, and participants

We conducted a retrospective single-center cohort study on sample data that were collected prospectively and validated. The study was performed at UHS, a tertiary-care and teaching hospital. In this study, we focused on two key groups in ADE detection: women and the elderly. Data from the UHS were collected from patients admitted in 2007 or 2011 to any ward of the Department of Internal Medicine (I+II). All female adult patients (age ≥ 18 years) admitted for at least 24 hours were eligible for the study. An additional age constraint was imposed on patients admitted in 2007: all of them had to be older than 75 years.

Data management and sample size

Demographic patient information, as well as clinical and laboratory values, were obtained through systematic interrogation and sampling of the hospital information system (HIS). A total of 70 patient cases were selected for the study; 22 from 2007 and 48 from 2011.

Data sources

Patient data were collected from various sources, such as the UHS’s HIS, or manually entered data. The following six main categories were used:

- Demographic data including demographic information such as age, sex, weight, height, pregnancy, or epidemiological studies.
- Laboratory findings provided by the HIS, such as serum creatinine, potassium, sodium, etc. Different time frames exist for absolute and relative findings. Absolute values are only taken into account within a time frame of three days prior to the data of

calculation, whereas relative values permit a time frame of seven days.

- Symptoms that occurred during the preceding seven days are integrated into the analysis.
- Diagnoses are defined according to the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10). A single diagnosis is identified by a 3- to 7-digit code (such as E87.5 for hyperkalemia). When more than one diagnosis from a diagnostic group is detected in the patient's chart, all of the diagnoses from this group are counted as a single trigger.
- Medications are specified by the Anatomical Therapeutic Chemical Classification System (ATC), according to which a single substance is defined by a 7-digit code.
- Hospital events denote any consultations of psychiatrists or trauma surgeons and internal accident reports during the hospital stay.

Risk score calculation

Risk score values and rules were determined by clinical and pharmacological experts and are assigned to ADE triggers which are processed by an algorithm to calculate an overall ADE risk score on a given scale (1–5). Based on the ADE risk score, appropriate reminders are sent to the physicians and pharmacists. Furthermore, reporting forms are prepopulated with the relevant patient data and the suspicion of an ADE.

The ADE risk score calculation consists of five main steps:

1. Patient data filtering. Only those data elements within a specified timeframe related to the calculation date and specified conditions are relevant for the calculation process. The time frames are based on clinical experience. Patient data shall be integrated when the medication that may cause an ADE has been administered during the preceding three days.
2. The recognition of at least one medication which may cause the possible ADE is a prerequisite for the calculation of an ADE risk score and the specific rules.
3. Depending on the number of positive triggers from each category, a contribution to the ADE risk score is calculated. The maximum value of all categories is added to the ADE risk score (see Table 1).
4. The ADE risk score is adapted by a value that depends on the quantity of the patient's medication (see Table 2).
5. Standardization of the last ADE risk score is the last calculation step. The maximum value for the ADE risk score is 5.

Table 1– The adverse drug event risk score increases, depending on the number of positive triggers

For each category
1–2 triggers with an ADE risk score of 1: increase value by 1
≥3 triggers with an ADE risk score of 1: increase value by 2
≥1 trigger with an ADE risk score of 2: increase value by 2
≥1 trigger with an ADE risk score of 3: increase value by 3

$$\text{new ADE risk score} = \text{old ADE risk score} + (\text{maximum of categories increase values})$$

Note: ADE, adverse drug event.

Table 2– The adverse drug event score increases, depending on the number of administered medications

The occurrence of 2–4 different medications from the medication lists causing the ADE for the use case results in an increase of the ADE risk score by 1.

The occurrence of >4 different medications from the medication lists causing the ADE for the use case results in an increase of the ADE risk score by 2.

Note: ADE, adverse drug event.

Knowledge base and data processing

Four highly critical clinical situations, namely hyperkalemia, hyponatremia, renal failure, and over-anticoagulation, were defined as use cases for the iMedication project. These constitute significant ADEs in internal and geriatric medicine. The four knowledge bases in iMedication are based on these use cases, which are implemented in Arden Syntax, which is a knowledge representation and processing language supported by HL7 International [15]. Each knowledge base consists of several medical logic modules (MLMs) [16, 17], which are the basic knowledge representation and processing units in Arden Syntax and are executed by an Arden Syntax engine [18]. In all there are 33 MLMs, taking 51 ADE triggers into account.

Data are processed as follows: First, all relevant data are collected and aggregated into an information block – a patient object – and forwarded to the Arden Syntax server. The Arden Syntax engine processes the obtained information and returns one result object for each knowledge base and each day. The return objects contain complete patient data, thus permitting the explanation and tracing of decisions made by the iMedication system. Also, for each category, the fired triggers are stored and attached. The result object includes information on the severity of the detected ADE; this information determines the ADE risk score.

Presentation of results

We use patient demographic information (age, length of stay, number of verified ADEs) and treatment information (number of administered medications) to describe the patient population. We also discuss the number and risk score of ADE triggers during the study period. We define each ADE trigger with a risk score ≥ 4 as a “positive test”, and ADEs with lower risk scores as a “negative test”. Using this classification, we determine the effectiveness of the system as well as its sensitivity (SEN), specificity (SPE), positive predictive value (PPV) and negative predictive value (NPV) metrics.

Results

The mean age of the patients was 76.5 years (standard deviation 13.3 years, minimum 43 years, maximum 99 years). The mean duration of the hospital stay was 11.2 days (standard deviation 10 days, minimum 2 days, maximum 53 days). On average, a patient received 8.5 medications during his/her stay (standard deviation 4.7; minimum 1 medication, maximum 29 medications).

Of the 70 patients included in the study, 16 (22.8%) experienced one or several ADEs confirmed by patient caregivers. Twelve patients with ADEs were registered in the study population of 2007, and 4 patients with ADEs among those examined in 2011. In all 26 ADEs were confirmed for the four medical situations implemented in the knowledge base: 2 for hyperkalemia, 13 for hyponatremia, 8 for renal failure, and 3 for over-anticoagulation.

A total of 428 triggers were generated during the study period. An overview of these triggers and their scores are shown in Table 3.

Table 3— Number of triggers generated during the study period and their associated scores.

Trigger score	Frequency
Score 0	306
Score 1	9
Score 2	9
Score 3	34
Score 4	34
Score ≥ 5	36
Total	428

Using the previously mentioned classification for a “positive test” and a “negative test”, we constructed a 2x2 contingency table (Table 4). Based on the absolute numbers in the contingency table, the system showed a SEN of 85%, a SPE of 88%, a PPV of 31%, and a NPV of 99%.

Table 4— 2x2 Contingency table for the study results

	ADE confirmed	ADE absent	Total
Positive test	22	48	70
Negative test	4	354	358
Total	26	402	428

Note: ADE, adverse drug event.

Discussion

We present the iMedication system, a computerized system that supports pharmacovigilance by detecting and reporting potential ADEs. We outlined the underlying principles and mechanics of the system, and established the sensitivity and specificity of the current pilot system. However, it needs to be refined before it can qualify as a trustworthy alarming system (PPV 31%).

Computerized trigger tools for inpatient ADEs perform moderately well, are inexpensive to use, and already deployed in many hospitals [19]. The iMedication system was able to correctly identify 85% of all ADEs, which is many times higher than the numbers of commonly reported ADEs (1–13%) [4]. According to a recent study, only 4.5–5.5% of ADEs are reported in Austria [21].

The iMedication system is able to help clinicians in many ways. First, a retrospective evaluation of clinical data permits quality assurance through statistical analysis of detected potential ADEs. Second, physicians are given active feedback (notifications) during the treatment of their patients, thus enabling them to take corrective measures in a timely manner.

Finally, the iMedication system supports (semi-)automated ADE reporting by notifications to the pharmacist with prepopulated forms. As a result, ADEs can be avoided or corrected. When they do occur, their reporting consumes less resources.

The limitations of the study are worthy of mention. First, in the present evaluation phase, data input is accomplished semi-automatically because all relevant patient data are not available in electronic form. Furthermore, the four use cases currently implemented in the knowledge base need to be evaluated in a wider setting and improved in order to avoid alert fatigue. Finally, additional studies will be carried out to evaluate the iMedication phenomenon of much more frequent ADE reports to the AGES than is achieved by conventional reporting.

Conclusion

We showed that a comprehensive solution for the (semi-) automated detection and reporting of ADEs is not only feasible but also effective. Given the fact that the tracking and reporting of ADEs occur on a voluntary basis, the integration of an automated computerized method in clinical routine would provide more information about the scope of the ADE problem at a minimal expense of resources.

Acknowledgements

The iMedication project is funded by Research, Innovation, Technology – Information Technology (FIT-IT [20]), an initiative of the Austrian Federal Ministry of Transport, Innovation, and Technology. The members of this consortial project are Salzburg Research Forschungsgesellschaft mbH, Gemeinnützige Salzburger Landeskliniken Betriebsgesellschaft mbH, Paracelsus Medical Private University Salzburg, Landesapothek am St. Johannis-Spital Salzburg, and Medexter Healthcare GmbH.

References

- [1] L.L. Leape, Error in Medicine, *JAMA* **272** (1994), 1851–1857.
- [2] M.A. Makary and M. Daniel, Medical Error—The Third Leading Cause of Death in the US, *BMJ* **353** (2016), i2139.
- [3] F.A. Griffin and R.K. Resar, *IHI Global Trigger Tool for Measuring Adverse Events (Second Edition)*, IHI Innovation Series White Paper, Institute for Healthcare Improvement, Cambridge, MA., 2009. Available from <http://www.IHI.org>, last access: 17 December 2016.
- [4] A. Zafar, J. Hickner, W. Pace, and W. Tierney, An Adverse Drug Event and Medication Error Reporting System for Ambulatory Care (MEADERS), *AMIA Annu Symp Proc 2008* (2008), 839–843.
- [5] A. Bobb, K. Gleason, M. Husch, J. Feinglass, P.R. Yarnold, and G.A. Noskin, The Epidemiology of Prescribing Errors: The Potential Impact of Computerized Prescriber Order Entry, *Arch Intern Med* **164** (2004), 785–792.
- [6] J.S. de Bruin, W. Seeling, and C. Schuh, Data Use and Effectiveness in Electronic Surveillance of Healthcare Associated Infections in the 21st Century: A Systematic Review, *J Am Med Inform Assoc* **21** (2014), 942–951.
- [7] A.J. Forster, H.J. Murff, J.F. Peterson, T.K. Gandhi, and D.W. Bates, The Incidence and Severity of Adverse

- Events Affecting Patients after Discharge from the Hospital, *Ann Intern Med* **138** (2003), 161-167.
- [8] Definition of pharmacovigilance (WHO, 2002)
- [9] E. Chazard, C. Preda, B. Merlin, G. Ficheur, and R. Beuscart, Data-Mining-Based Detection of Adverse Drug Events, *Stud Health Technol Inform* **150** (2009), 552-556.
- [10] E. Chazard, A. Baceanu, R. Marcilly, S. Bernonville, G. Ficheur, and R. Beuscart, A Web Tool for Automated Adverse Drug Events Detection: The ADE Scorecards, *Stud Health Technol Inform* **169** (2011), 165-169.
- [11] F. Cao, X. Sun, X. Wang, B. Li, J. Li, and Y. Pan, Ontology-Based Knowledge Management for Personalized Adverse Drug Events Detection, *Stud Health Technol Inform* **169** (2011), 699-703.
- [12] Snomed, <http://www.ihtsdo.org/snomed-ct/>, last access: 17 December 2016.
- [13] J.D. Duke and J. Friedlin, ADESSA: A Real-Time Decision Support Service for Delivery of Semantically Coded Adverse Drug Event Data, *AMIA Annu Symp Proc 2010* (2010), 177-181.
- [14] T. Morimoto, T.K. Gandhi, A.C. Seger, T.C. Hsieh, and D.W. Bates, Adverse Drug Events and Medication Errors: Detection and Classification Methods, *Qual Saf Health Care* **13** (2004), 306-314.
- [15] Health Level Seven, *Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10)*, 2015. Available at: http://www.hl7.org/implementation/standards/product_brief.cfm?product_id=372, last access: 8 November 2016.
- [16] K.-P. Adlassnig and A. Rappelsberger, Medical Knowledge Packages and Their Integration into Health-Care Information Systems and the World Wide Web, *Stud Health Technol Inform* **136** (2008), 121-126.
- [17] G. Hripcsak, Writing Arden Syntax Medical Logic Modules, *Comput Biol Med* **24** (1994), 331-363.
- [18] K.-P. Adlassnig and K. Fehre, Service-Oriented Fuzzy-Arden-Syntax-Based Clinical Decision Support, *Indian Journal of Medical Informatics* **8** (2014), 75-79.
- [19] H.J. Mull, J.R. Nebeker, S.L. Shimada, H.M. Kaafarani, P.E. Rivard, and A.K. Rosen, Consensus Building for Development of Outpatient Adverse Drug Event Triggers, *J Patient Saf* **7** (2011), 66-71.
- [20] FIT-IT, <http://www.fit-it.at>, last access: 10 April 2016.
- [21] MIM-Project Drug safety 3 – Analysis of Adverse Drug Events in Austrian hospital discharge diagnoses between 2001 and 2011 (ADE3), <https://cemsis.meduniwien.ac.at/mim/wf/projekte/drug-safety3/>, last access: 10 April 2017.

Address for correspondence

Univ.-Prof. DI Dr. Klaus-Peter Adlassnig, FACMI Section for Artificial Intelligence and Decision Support
Center for Medical Statistics, Informatics, and Intelligent Systems
Medical University of Vienna
Spitalgasse 23, 1090 Vienna, Austria
E-mail: klaus-peter.adlassnig@meduniwien.ac.at and
Medexter Healthcare GmbH Borschkegasse 7/5, 1090 Vienna, Austria
E-mail: kpa@medexter.com

Usability Evaluation of an EHR's Clinical Notes Interface from the Perspective of Attending and Resident Physicians: An Exploratory Study

Rubina F. Rizvi^a, Jenna L. Marquard^b, Marcus A. Seywerd^c, Terrence J. Adam^{a,d}, Jed T. Ellison^e,
Gretchen M. Hultman^a, Kathleen A. Harder^{a,f}, Genevieve B. Melton^{a,g}

^a Institute for Health Informatics, University of Minnesota, Minneapolis, MN, USA

^b Department of Industrial Engineering, University of Massachusetts, Amherst, MA, USA

^c EPIC Innovations, Fairview Health Services, University of Minnesota, Minneapolis, MN, USA

^d College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

^e Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN, USA

^f Center for Design in Health, College of Design, University of Minnesota, Minneapolis, MN, USA

^g Department of Surgery, University of Minnesota, Minneapolis, MN, USA

Abstract

Usability gaps between current and future improved Electronic Health Record (EHR) system designs exist due to insufficient incorporation of User-Centered Design (UCD) principles during System Development Life Cycle (SDLC). Usability of a commercial, inpatient EHR clinical notes documentation interface was analyzed from standpoints of two provider groups employing two standardized patient cases. Both objective and subjective data were collected from attending (n=6) and resident physicians (n=8) through usability testing employing a mixed method approach. The study results suggested that (i) EHR usability and desirability is influenced by user characteristics, (ii) workloads associated with H&P and progress notes writing are perceived differently between two groups, (iii) repeated task performance improves user efficiency and (iv) user performance is correlated to their subjective system assessments. Understanding usability of clinical documentation interface from perspectives of two different user groups, provides interface designers with an opportunity to develop an EHR system centered on UCD principles.

Keywords:

Electronic Health Record (EHR); User-Computer Interface; Documentation

Introduction

While Electronic Health Record (EHR) systems have been widely adopted with the ultimate goal of improved health care delivery [1], substantial gaps exist between the current state of EHRs and their potential usefulness [2]. Poor EHR usability appears to be a major factor for this discrepancy [2]. To facilitate optimal end product usability, it is critical to understand end users' "usage behavior", considered a core feature of a User-Centered Design (UCD) approach [3,4]. The UCD philosophy is that "the final product should suit the users, rather than making the users suit the product" [5]. According to the International Organization for Standardization (ISO)-framework used in this research study, usability is defined as the, "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [6]. Similarly, in EHR design, user involvement throughout the System Development Life Cycle (SDLC) can facilitate the development of

systems that are easy to learn and remember, efficient, minimize errors and improve user satisfaction [7], which could improve EHR adoption and better patient outcomes [8].

Despite the critical role of the Human Computer Interaction (HCI) in the SDLC process [9], it is often neglected during EHR interface design. Usability studies on EHRs' clinical decision support system and user interfaces for medical equipment have been done in the past [10,11], but there are not many studies focusing on clinical notes documentation within an EHR interface [12-15], with only few studies done on usability evaluation and prototyping of clinical notes user interfaces in the medical domain [12-17]. Similarly, usability of a system could vary with vendor types and user profiles (e.g. clinical experience, EHR training, age, gender, technology skills etc.) However, few research studies incorporate usability comparisons from diverse user perspectives (e.g., expert users vs. novice users; physician vs. patients; users vs. usability experts) [18-20].

Usability testing is accepted as the most effective usability methodology with greatest strategic impact [21]. It is an "activity that focuses on observing users working with a product and performing tasks that are real and meaningful to them" [9]. The purpose of this study is to quantify EHR usability around inpatient notes usage focusing on the clinical note documentation and clinical note viewing interface, an area that poses tremendous challenges to physicians and other clinicians working under time limitations [13]. Both objective and subjective data on users' task performance were collected from two user groups (i.e., attendings and residents) and analyzed via usability metrics as defined by ISO (i.e., effectiveness, efficiency and satisfaction) [6]. Supplementary data were also analyzed for subjective workload using the NASA-TLX instrument [22] and system desirability with Product Reaction Cards (PRC) [23]. The insight gained through this research provides an opportunity to better understand EHR usability around clinical documentation from the standpoints of two provider groups and identify usability gaps to benchmark future EHR design.

Methods

This research study evaluated the usability of an enterprise EHR (Epic Systems Corporation) system at Fairview Health

Services, University of Minnesota Medical Center (UMMC). The study specifically focused on clinical documentation tasks (e.g., H&P and progress note writing). Scenario-based usability testing was conducted on two high fidelity simulated test patient charts [24] in an Epic test environment replicating the real work environment, both in design and functionality. Testing was done at the usability laboratory.

Study sample

Physician participants (n=14) were from two user groups: attendings (n=6) and residents, excluding interns (n=8). Participants were in all cases either trained in Internal medicine or Family medicine with past and/or current inpatient experience with the Epic Fairview EHR. Detailed user characteristics categorized by user group, are summarized in Table 1.

Table 1- users characteristics

	Age (M/F)	Clinical Exp. (Yrs.)	Technology Exp.	Epic Exp. (Yrs.)	Epic proficiency
A	31 (F)	≤5	Somewhat	5-10	Average
A	43 (F)	> 10	Less	5-10	Average
A	xx (F)	≤5	Somewhat	5-10	Average
A	43 (F)	> 10	Very	> 10	Proficient
A	36 (M)	≤5	Somewhat	< 5	Proficient
A	39 (M)	≤ 5	Somewhat	5-10	Average
R	30 (F)	≤ 5	Somewhat	< 5	Average
R	xx (F)	≤ 5	Somewhat	< 5	Average
R	xx (M)	≤ 5	Somewhat	< 5	Average
R	30 (F)	≤ 5	Somewhat	< 5	Average
R	29 (M)	≤ 5	Somewhat	< 5	Proficient
R	26 (M)	≤ 5	Very	< 5	Proficient
R	29 (M)	≤ 5	Somewhat	< 5	Proficient
R	29 (F)	≤ 5	Somewhat	< 5	Average

A=Attendings; R=Residents; Clinical Exp.: Clinical Experience (Residency training and later); Epic Exp.: Total years of experience with using Epic

Participation was voluntary and participants received \$50/hour. Each session was 2.5-3 hours long and each physician was at least 24 hours off night call on their day of data collection. The study protocol was approved by the Institutional Board Review.

Data collection and analysis

Two simulated, high fidelity test patient charts with rich, realistic clinical data were created in an Epic test environment to provide scenario-based EHR usability testing [24]. Patients were built from representative cases after extensive discussion among five experts: the lead EHR physician trainer (MS) and four physician informaticists (RR, TA, GMM & EA). Patient cases with similar complexities were selected using a Charlson weighted comorbidity index and number of prior admissions, clinic visits, and clinical notes. In both clinical scenarios, patients with a history of Chronic Obstructive Pulmonary Disease and Congestive Heart Failure presented in the emergency department with sudden onset shortness of breath. Each participant was assigned two patient cases in a random order employing an online randomization tool [25]. A Randomized blocked design approach was used to create balanced distribution of test patients across two groups. Each participant per-

formed the same tasks of entering a H&P and a day 1 progress note, on each test patient's chart.

Raw data was extracted employing Tobii studio version 3.4.5 and was evaluated in three ways: (a) user satisfaction, via the System Usability Scale (SUS) questionnaire [26,27] (b) efficiency, via time on tasks, key presses, & mouse clicks and (c) effectiveness, via note quality using the Physician Documentation Quality Instrument-9 (PDQI-9) [28] and overall Gestalt judgment [29]. Data from each user group was also analyzed for subjective workload index using the NASA-TLX questionnaire [22] and system desirability via Product Reaction Cards (PRC) listing 118 words [23]. All participants were asked to circle their top 5 choices, which were later compiled as a word cloud and Venn diagram to visualize total and unique word selection by each user group.

Note quality assessment was performed by two co-authors/physicians (RR and TA) using standardized metrics as previously reported with the Physician Documentation Quality Instrument-9 (PDQI-9) [28] and overall Gestalt judgment [29]. Pretesting of these instruments for note quality assessment was conducted on a set of unrelated notes to ensure that both reviewers shared a common understanding of item scoring. Once consensus was achieved, both evaluators reviewed and assessed approximately 14% of notes (8 of 56 notes). The consistency in quality assessment was checked by inter-coder agreement with final mean agreement for PDQI-9 of 81% (kappa=0.69) and Gestalt scoring of 87.5% (kappa=0.71). We report summative statistics using SAS enterprise guide 5.1 and StatPlus LE 6.0.3 (a statistical software plugin for Macintosh) with means and standard deviation (sd).

Results

While not statistically different, user satisfaction with respect to overall usability of clinical note entry was perceived worse by attendings (mean SUS = 60.8 ± 15.6 (i.e., marginal usability)) compared to residents (mean SUS = 73.4 ± 13.5, (i.e., acceptable usability)), despite longer average Epic experience among attendings (≥ 5 years, n=5/6) compared to residents (< 5 years, n=8/8). The SUS and their interpretation [27] are illustrated in Fig 1.

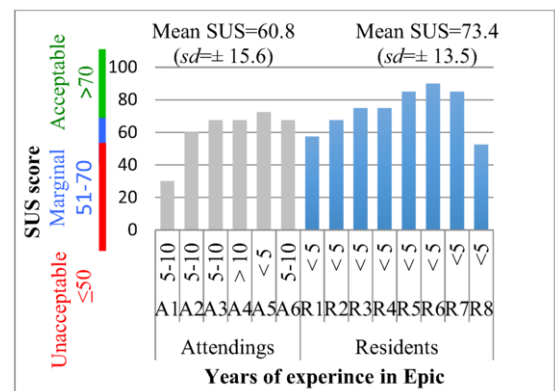


Figure 1-SUS based on users' characteristics

Efficiency was quantified based on time on task, key presses, and mouse clicks. H&P writing was more time-intensive than progress notes for both attendings (26.2 ± 9.7 vs. 14.0 ± 6.4 minutes) and residents (24.2 ± 7.7 vs. 12.3 ± 4.5 minutes). Residents took slightly less time than attendings writing both

H&P (24.2 ± 7.7 vs. 26.2 ± 9.7 minutes) and progress notes (12.3 ± 4.5 vs. 14.0 ± 6.4 minutes). Time on task decreased from the 1st to 2nd patient, except for progress note writing among residents (Fig 2).

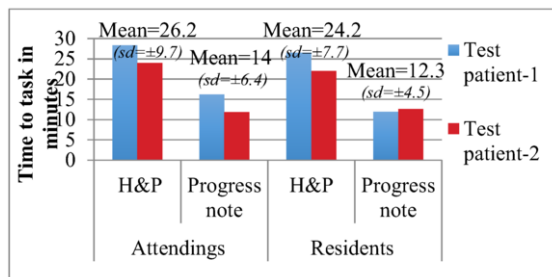


Figure 2-Time to tasks comparison between two user groups

More key presses (KP) and mouse clicks (MC) were observed with H&P as compared to progress note writing for both attendings (KP=2,644 ± 1535 vs. 1,433 ± 682, MC=201 ± 83 vs. 126 ± 60) and residents (KP=3,468 ± 1,199 vs. 1,758 ± 689, MC=214 ± 82 vs. 112 ± 46) with residents generally performing more key presses and mouse clicks compared to attendings with exception of progress notes where attendings had more mouse clicks. The number of key presses and mouse clicks decreased from the 1st to 2nd patient, except for residents' number of mouse clicks during progress note writing (Fig 3, 4).

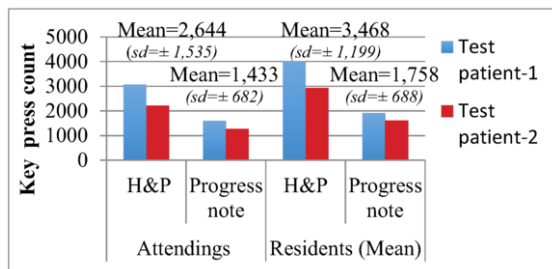


Figure 3-Number of key presses comparison between two user groups

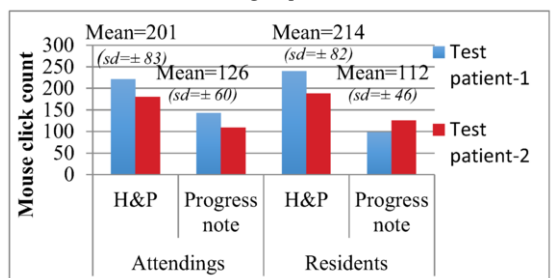


Figure 4-Number of mouse clicks comparison between two user groups

Effectiveness, as measured through PDQI-9 scores on note quality showed no quality differences between H&P and progress notes by attendings (34.9 ± 3.8 vs. 34.8 ± 4.8), though resident progress notes were slightly higher quality than H&P notes (35.5 ± 6.3 vs. 33.8 ± 4.0). Attendings' H&P notes (34.9 ± 3.8 vs. 33.8 ± 4.0), and residents' progress notes (35.5 ± 6.3 vs. 34.8 ± 4.8) showed only minimal quality differences. No noticeable differences in note quality between attending and residents were detected through Gestalt scoring both for H&P (3.7 ± 0.7 vs. 3.8 ± 0.8) and progress notes (3.9 ± 0.9 vs. 4.0 ± 1.0). PDQI-9 scores increased from the 1st to 2nd patient, except for residents' progress notes (Fig 5).

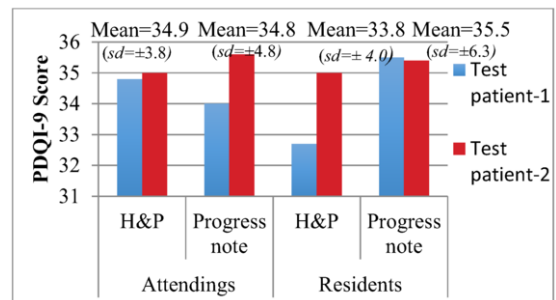


Figure 5-PDQI-9 scores comparison between two user groups

The NASA-RTLX questionnaire revealed that H&P note writing had higher overall workload (OW) than progress note writing among both attendings (27.8 ± 11.4 vs. 27.2 ± 16.0) and residents (33.6 ± 16.7 vs. 22.5 ± 10.2). Residents also had considerably higher subjective OW for H&P note writing (33.6 ± 16.7 vs. 27.8 ± 11.4), while attendings had higher subjective OW for progress note writing (27.2 ± 16.0 vs. 22.5 ± 10.2). There was no effect of patient order on perceived workload (Fig 6)

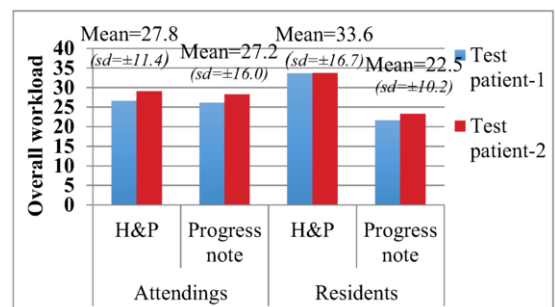


Figure 6-NASA-RTLX scores comparison between two user groups

Pearson correlation coefficient calculations were performed on the data after visually inspecting the distributions for normality. The results showed some correlation between metrics: NASA & SUS (-0.79 vs. -0.55 i.e., strong negative), NASA & Gestalt (-0.27 vs. -0.26 i.e., fair degree of negative), PDQI-9 & Gestalt (0.82 vs. 0.70 i.e., strong positive), Gestalt & SUS (0.39 vs. 0.30 i.e., fair degree of positive). Other metrics showed weak or no correlation: PDQI-9 & SUS (0.13 vs. 0.14), time on task & PDQI-9 (-0.18 vs. 0.10), and time on task & Gestalt (0.23 vs. 0.20).

System desirability analysis compared the proportion of positive vs. negative terms from a comprehensive list of 118 words (23). A higher percentage of positive as compared to negative words were selected both by attendings (63% vs. 37%) and residents (73% vs. 28%). Attendings selected a higher percentage of negative words (37% vs. 28%) while residents selected a higher percentage of positive words (73% vs. 63%) words respectively as depicted in the word cloud images (Fig 7). Similar results were seen for unique word selection as shown in the Venn diagram.

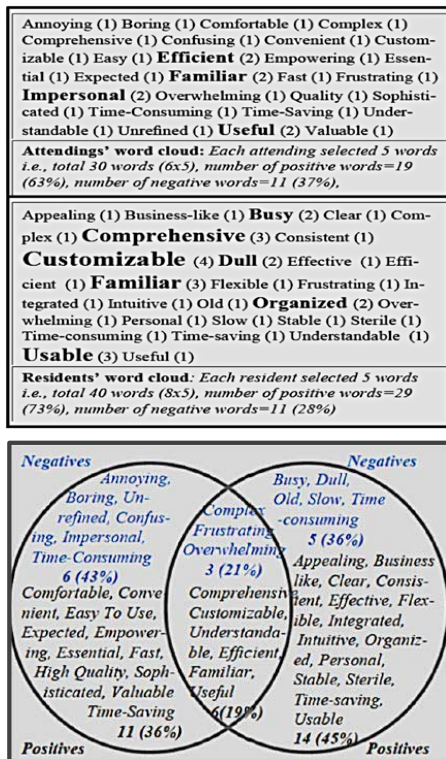


Figure 7-Product Reaction Cards- Attendings and residents

Discussion

This research study is an important initial step towards understanding the usability of EHR clinical notes documentation from attending and resident physician perspectives. EHR usability, as quantified through objective measures of user performance and their subjective perceptions, varied with each group, note type, and repeated tasks. Varying degrees of correlation were also discovered between variables, suggesting that user performance is related to their subjective system assessments. The insight gained through this research provides an opportunity to better understand EHR clinical documentation usability, identify and address existing usability gaps, and establish benchmarks for future EHRs.

Based on the SUS, residents perceived the system to have "acceptable usability" while attendings perceived the system to have "marginal usability", despite attendings having more experience with Epic. Similarly, system desirability was considered better among residents compared to attendings, with a higher percentage of positive words used to describe the system. Since residents are generally exposed to EHRs early in their medical training and tend to have little exposure to traditional paper charting, this may explain more favorable responses to EHR usability and desirability. Additionally, resident, predominantly males, tended to be younger, and rated themselves as having more technical experience, leading to easier technology adoption. Thus, user characteristics appear to be a critical factor for EHR usability.

In terms of efficiency, as quantified by time on task, key presses and mouse clicks, attendings and residents both took significantly more effort with H&P compared to progress note writing. Residents perceived less subjective workload associated with progress notes suggesting that residents were more at

ease in writing progress notes. A potential reason for this is the nature of progress note writing task itself, which is more repetitive and most likely to be influenced by a system's usability (e.g., copying and pasting, auto population, multiple screen panel functionalities, etc.) In comparison, attendings showed less subjective workload with H&P writing suggesting that they are better skilled in writing H&P notes, a cognitively demanding task which involves providing a reason for admission and providing initial patient management direction. Thus, targeted note documentation training of physicians where there is a lack of proficiency (e.g., H&P among residents and progress note in attendings), would be a reasonable approach to consider. No noticeable difference in note quality between attending and residents was detected through Gestalt scoring.

Generally, efficiency improved as users performed the same note writing tasks on the 2nd patient with the exception of progress note writing among residents. The plausible explanation of the observed differences may be due to user familiarity with the system and faster cognitive processing as a result of repeated task performance, as well as specifics around the second patient case. No effect of patient order was observed on perceived workload while there was some indication of improvement in note quality, especially progress notes, among attendings and H&P writing among residents.

We discovered that increases in subjective workload (NASA) were associated with decreases in user satisfaction (SUS) and note quality (per Gestalt). Higher satisfaction was associated with better quality notes (per Gestalt). We found a strong positive correlation between PDQI-9 & Gestalt, but no correlation was detected between PDQI-9 & SUS or with time on task and note quality for both PDQI-9 & Gestalt.

There are some limitations associated with this study, including a small sample size lacking significant inferential statistical results. Generalizability is limited due to the inclusion of physicians (MDs) with training in either Internal medicine or Family medicine and testing of inpatient EHR interfaces only. Future studies are mandated with larger sample size, more diverse group of participants (e.g., medical students, nurses, ER physicians etc.) and on different EHR systems (e.g., EHRs used in ambulatory settings, emergency rooms and as well as specialty specific EHRs). Additionally, the impact of other user characteristics (e.g., age, gender, etc.) needs to be explored further. There are also some limitations associated with usability testing itself (e.g., relevance of scenarios/ tasks being tested, system speed and connectivity etc.). In addition, understanding physicians' EHR usage behaviors around clinical note documentation, the focus of our next study, is an essential area that needs to be further explored.

Conclusions

We discovered that EHR usability measures of satisfaction, efficiency, and effectiveness vary with users' characteristics, specific note types, and from repeated performance of the same task on consecutive patients' charts. This study provides preliminary, yet essential information on objective measures of user performance and their perceptions of EHR usability around clinical notes usage. These measures can serve as initial guidance to build EHR interfaces grounded on a "User-Centered Design" approach.

Acknowledgements

We would like to thank the staff at Center for Neurobehavioral Development for collaborating, Elliot G. Arsoniadis, all resi-

dents and attendings and Fairview Health Services. This work was supported by National Science Foundation (NSF) Award #CMMI-1150057 (JM) and the Agency for Healthcare Research and Quality (AHRQ) Award #R01HS022085 (GM). The content is solely the responsibility of the authors and does not represent the official views of the NSF or AHRQ

References

- [1] Hillestad R, Bigelow J, Bower A, Girosi F, Meili R, Scoville R, et al. Can electronic medical record systems transform health care? Potential health benefits, savings, and costs. *Health Aff (Millwood)* 2005 Sep-Oct;24(5):1103-1117.
- [2] Zhang J, Walji MF. TURF: Toward a unified framework of EHR usability. *J Biomed Inform* 2011 12;44(6):1056-1067.
- [3] Norman DA, Draper SW. User centered system design; new perspectives on human-computer interaction: L. Erlbaum Associates Inc.; 1986.
- [4] Zheng K, Padman R, Johnson M, Diamond H. An interface-driven analysis of user interactions with an electronic health records system. *J Am Med Inform Assoc* 2009 Mar-Apr;16 (2):228-237.
- [5] Baxter K. Understanding your users: a practical guide to user research methods. Second edition. Waltham, MA : Morgan Kaufmann; 2015.
- [6] Abran A, Khelifi A, Suryan W, Seffah A. Usability meanings and interpretations in ISO standards. *Software Quality Journal* 2003;11(4):325-338.
- [7] Nielsen J. Usability engineering. : Elsevier; 1994.
- [8] Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. *J Biomed Inform* 2005 2;38(1):75-87.
- [9] Barnum CM, Barnum CM. Usability testing essentials ready, set-- test! Burlington, Mas. : Morgan Kaufmann; 2010.
- [10] Carroll C, Marsden P, Soden P, Naylor E, New J, Dornan T. Involving users in the design and usability evaluation of a clinical decision support system. *Comput Methods Programs Biomed* 2002;69(2):123-135.
- [11] Liu Y, Osvalder A. Usability evaluation of a GUI prototype for a ventilator machine. *J Clin Monit Comput* 2004;18(5-6):365-372.
- [12] Sharda P, Das AK, Cohen TA, Patel V. Customizing clinical narratives for the electronic medical record interface using cognitive methods. *Int J Med Inf* 2006;75(5):346-368.
- [13] Rizvi RF, Harder KA, Hultman GM, Adam TJ, Kim M, Pakhomov SV, et al. A comparative observational study of inpatient clinical note-entry and reading/retrieval styles adopted by physicians. *Int J Med Inf* 2016;90:1-11.
- [14] Brown P, Marquard J, Amster B, Romoser M, Friderici J, Goff S, et al. What Do Physicians Read (and Ignore) in Electronic Progress Notes? *Appl Clin Inform* 2014;5(2):430-444.
- [15] Using High-Fidelity Simulation and Eye Tracking to Characterize EHR Workflow Patterns among Hospital Physicians. *AMIA Annual Symposium Proceedings*: 2015.
- [16] Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *J Am Med Inform Assoc* 2006 May-Jun;13(3):277-288.
- [17] Brown SH, Hardenbrook S, Herrick L, St Onge J, Bailey K, Elkin PL. Usability evaluation of the progress note construction set. *Proc AMIA Symp* 2001:76-80.
- [18] Clarke MA, Belden JL, Kim MS. Determining differences in user performance between expert and novice primary care doctors when using an electronic health record (EHR). *J Eval Clin Pract* 2014;20(6):1153-1161.
- [19] Ozok AA, Wu H, Garrido M, Pronovost PJ, Gurses AP. Usability and perceived usefulness of personal health records for preventive health care: A case study focusing on patients' and primary care providers' perspectives. *Appl Ergon* 2014;45(3):613-628.
- [20] Corrao NJ, Robinson AG, Swiernik MA, Naeim A. Importance of testing for usability when selecting and implementing an electronic health or medical record system. *J Oncol Pract* 2010 May;6(3):120-124.
- [21] A toolkit for strategic usability: results from workshops, panels, and surveys. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*: ACM; 2000.
- [22] NASA-task load index [NASA-TLX]; 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*: Sage Publications; 2006.
- [23] Benedek J, Miner T. Measuring Desirability: New methods for evaluating desirability in a usability lab setting. *Proceedings of Usability Professionals Association* 2002;2003:8-12.
- [24] Leveraging a Proprietary Electronic Health Record System to Create High-Fidelity, Smart Test Patients. *iHealth 2016 Clinical Informatics Conference*; 2016; ; 2016.
- [25] Urbaniak GC, Plous S. Research Randomizer (Version 4.0) [Computer software].
- [26] Brooke J. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 1996;189(194):4-7.
- [27] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 2009;4(3):114-123.
- [28] Stetson PD, Bakken S, Wrenn JO, Siegler EL. Assessing Electronic Note Quality Using the Physician Documentation Quality Instrument (PDQI-9). *Appl Clin Inform* 2012;3(2):164-174.
- [29] Penaloza A, Verschuren F, Meyer G, Quentin-G S, Soulie C, Thys F, et al. Comparison of the unstructured clinician gestalt, the wells score, and the revised Geneva score to estimate pretest probability for suspected pulmonary embolism. *Ann Emerg Med* 2013;62(2):117-124.

Address for correspondence

Genevieve B. Melton, MD, PhD
Institute for Health Informatics and Department of Surgery
420 Delaware Street SE, Mayo Mail Code 450
Minneapolis, MN 55455
Tel: + 612-625-7992, E-mail: gmelton@umn.edu

Strengthening Data Confidentiality and Integrity Protection in the Context of a Multi-Centric Information System Dedicated to Autism Spectrum Disorder

Mohamed Ben Said^a, Laurence Robel^b, Bernard Golse^b, Jean Philippe Jais^a

^aParis Descartes University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, APHP - Necker Enfants Malades Hospital, Paris, France

^bDepartment of Child Psychiatry, APHP - Necker Enfants Malades Hospital, Paris, France

Abstract

Autism spectrum disorders (ASD) are complex neuro-developmental disorders affecting children in early age. Diagnosis relies on multidisciplinary investigations, in psychiatry, neurology, genetics, electrophysiology, neuro-imaging, audiology, and ophthalmology. To support clinicians, researchers, and public health decision makers, we developed an information system dedicated to ASD, called TEDIS. It was designed to manage systematic, exhaustive and continuous multi-centric patient data collection via secured internet connections. TEDIS will be deployed in nine ASD expert assessment centers in Ile-DeFrance district. We present security policy and infrastructure developed in context of TEDIS to protect patient privacy and clinical information. TEDIS security policy was organized around governance, ethical and organisational chart-agreement, patients consents, controlled user access, patients' privacy protection, constrained patients' data access. Security infrastructure was enriched by further technical solutions to reinforce ASD patients' privacy protection. Solutions were tested on local secured intranet environment and showed fluid functionality with consistent, transparent and safe encrypting-decrypting results.

Keywords:

Autism Spectrum Disorder; Data Encryption; Security Policy

Introduction

Electronic health record and patient personal information protection are critical for trust and reliability of an information system whether it is used within a locally accessible health care facility or over the Cloud and Internet network. A security policy and infrastructure was developed for both confidentiality and integrity of information managed in TEDIS [1,2], an information system, dedicated to patients with Autism Spectrum Disorder (ASD) [3].

TEDIS was developed in the Department of Child Psychiatry and the Department of Biostatistics and Medical Informatics (SBIM) at Necker Hospital, in Paris, France. TEDIS is aimed at supporting medical community and public health decision-makers to improve knowledge about psychopathological, physio-pathological, and etiological processes involved in ASD, to follow-up incident patient cohorts and to support decision making, patient care, and research. To reach these objectives, TEDIS was designed to manage systematic, exhaustive and continuous multi-centric ASD patients data collection via secured internet connections. Nine ASD patients' assessment and multidisciplinary centers in Ile-de-

France district collect a large amount of valuable clinical data. Eight centers evaluate children from 3 years old to-up and one center provides ASD assessment information for teenagers and adults. ASD patient assessment lasts about three weeks in a specialized ASD assessment center. Patients undergo evaluation in a variety of domains: clinical assessment of psychological, motor and speech development, consultation in neurology and genetics, and specialized para-clinical exams including brain MRI with spectroscopy, contextual EEG, standard and high resolution caryotype, searches for chromosomes alterations by fluorescent in situ hybridization (FISH), Comparative Genomic Hybridization Array (CGH-Array) and for metabolic disorders. Additional investigations guided by clinical exam are prescribed as necessary. A written report is issued at the end of the assessment. It is transmitted to parents, care providers and/or institutions involved in the patient care. Formal demands are consequently addressed to specialized institutions to consider patient admission, to the Institution of Handicapped Persons¹(MDPH) to qualify the handicap and to the Social Security medical officer to qualify the long term character of the disease²(ALD) in order to benefit from the medical care cost coverage. Treatment consists of life term care: series of early and individually adapted measures in domains of education, behaviour, and psychology. Treatment compliance, may significantly improve relational capabilities and social interaction with some degree of autonomy and possibility of language acquisition and non-verbal communication [1].

The national ethics and computing authority called Commission Nationale Informatique et Libertés (CNIL), authorized use of TEDIS within a conditional-framework based on patient' consents, respect of privacy, and protecting personal information [2]. The main philosophy behind confidentiality and privacy protection was to limit without affecting the research objectives, risk of altering data integrity and risk of unauthorized access which may lead to data loss or to establish a link between patient personal information and clinical information. CNIL formally requested to restrict access to authorized professionals and to store information which may directly or indirectly lead to identifying patients' (e.g. last name, first name, date of birth and sex), only in a strongly encrypted form on the database server.

On the other hand, TEDIS users particularly psychiatrists and clinicians, requested ability to visualize in plain text form patient identifying information – last name, first name, sex, and date of birth. Main point being to help directly

¹ In french : Maison des Personnes Handicapée (MDPH)

² In french : Affection de Longue Durée (ALD)

disambiguate patient's identity, facilitate linking clinical data to appropriate patient and facilitate quality control. Therefore, patient phonetic search on the web page using a list of patients assessed within one clinical ASD assessment center is possible allowing access to ASD patient record for consulting and editing.

We obtained CNIL agreement in mid-2014. Medical experts, clinicians and interns in child psychiatry department at Necker hospital started testing TEDIS application, over a stand-alone computer-server managed by SBIM at Necker Hospital. The experimentation helped adapting and improving system performance and stabilizing conceptual data model. It allowed conducting proof of concept and demonstrated feasibility of proposed scenarios of use of TEDIS in the nine ASD expert assessment centers. TEDIS patient privacy protection became crucial in the perspective of opening internet access to professionals in the ASD assessment centers.

In the following, we will recap security policy adopted and describe in more details methods adopted in security infrastructure to reinforce patient privacy protection in TEDIS. We will then discuss strengths and limits of our approach and perspective of future improvements and optimization. To begin with, we look at literature feedback in relation with the raised questions.

Literature review

Several articles address electronic health record security and privacy protection in the context of electronic health care information systems. Two major axes, with related subtopics of interest to our study may be distinguished: Security policy and Security infrastructure. *"The security policy should clearly define the guideline for creating, accessing and maintaining the integrity of patient e-health data and the scope of accountability for each responsible party. A security infrastructure covers the issues of login authentication, cryptography, access control,..., audit,..., and disaster recover"* [4].

Securing connection between the client and the Server

A major part of the security policy relies on the hosting environment of the application and database servers. French legislation defined personal health data hosting criteria in decree n°2006-6 of Jan 4th 2006 [5]. All medical information systems in France, managing personal and health information via the internet must be hosted at an Accredited Internet Hosting Provider in conformity with the legislation.

In addition, securing the connections between client and server would be provided at the application and network levels. The application security infrastructure would support secure connections between a client and server through encryption and authentication. Additional authentication controls and verifications are performed to reduce opportunity for interception of messages and unauthorized access to the server. *"At the network level, access to the Application server would be restricted by the client's internet protocol (IP) address. Each site that wishes to use the service must submit the fully qualified domain name and static IP address of the machine that will run the client application. In addition, the server audits all access and requests and reports unusual activity to the system administrator"* [6].

Access Control

R.R. Schell insisted on the *"scientific and engineering rigor needed for a trustworthy system to defend the security of computers in three dimensions at the same time: mandatory access control (MAC) policy, protection against subversion, and verifiability."* [7]. Jayabalan and O'Daniel emphasized ISO

22600-1:2014 principles and guidelines in a large systematic literature review of access control for electronic health record systems to protect patient's privacy. ISO 22600-1:2014 *defines principles and specifies services needed for managing privileges and access control for communication and use of health information distributed across policy domain boundaries. It highlights information sharing between diverse health care entities... without compromising privacy and integrity of EHR* [8]. In similar projects managed at SBIM-Necker for renal disease and for rare diseases domains, we developed experience with Discretionary Access Control (DAC) to organize users' access to dedicated information systems [9,10]. DAC is *"based on the decision of the owner, to limit access to patient data based on the identity of subjects and/or groups to which they belong"* [4].

On another hand, Mandatory Access Control (MAC) [4] presents an interesting approach in that *"it assigns security labels ... and allows access only to entities ... with distinct levels of authorization or clearance. MAC controls are enforced by the operating system or security kernel* [5].

Confidentiality-Privacy protection

Cryptography is used for reinforcing patient privacy protection and data integrity. It is *"the study of Secret (crypto-) writing (-graphy) that is concealing the content of message from all except the sender and the receiver and to authenticate the correctness of message to the recipient"* [11]. Cryptography goals *"include privacy or confidentiality, data integrity, authentication and non-repudiation"* [4,11,12].

Chenglang addressing *"internal staff member accessing and disclosing patient privacy information"* suggested to *"encrypt privacy data before" storing it on the server database "using traditional algorithms, so that the data even if being disclosed are also difficult to be decrypted and understood"* [13].

In the present work we focused on the following technologies to reinforce privacy protection: symmetric cryptography [11, 12,14], hash coding [15], Sealed Objects [16] and serialization [17,18].

- Symmetric cryptography: *"In symmetric key cryptography the same key is used to encrypt and decrypt the data...The main problem with symmetric key algorithm is to exchange the secret key and cipher text between the sender and the receiver. It requires a secure channel"* [11]. The authors compared symmetric keys (DES, 3DES, AES, Blowfish) performances. They concluded that *"blowfish algorithm is superior to the other algorithms: DES, AES and Triple DES on the basis of key size and security. Blowfish algorithm runs faster than other symmetric key encryption algorithms and gives better performance in terms of encryption time, decryption time and throughput."* [11].
- Hash coding: *"Whenever it is invoked on the same object more than once during an execution of a Java application, the hashCode method must consistently return the same integer..."* [6,15].
- Sealed Objects: The SealedObject class in Java enables a programmer *"to create an object and protect its confidentiality with a cryptographic algorithm"* [16], while
- Serialization: *"to serialize an object means to convert its state to a byte stream so that the byte stream can be reverted back into a copy of the object"* e.g. "Object Serialization supports the encoding of objects and the objects reachable from them, into a stream of bytes". Serialized objects may be stored in a file, directory or in a database [17,18].

We developed a solution which combines a sequence of cipher processes of patient private information starting from the user interface, to the middleware and to the server database. Whenever requested, backward decipher processes are executed from the server database to the middleware and to the end-user interface.

We organized the encryption/decryption sequence presented in the literature review section to provide a stable, performant and robust solution for preserving data integrity, hiding private data on the server side data base.

Methods

Security policy

Hosting

In order to comply with the decree [5], we recently had a formal agreement with the informatics department of Assistance Publique des Hôpitaux de Paris (APHP), as an Accredited Hosting Provider in Health Personal Data, certified by the Public Health Ministry to host TEDIS application and database servers and offers secured internet connections to the professionals in the ASD assessment centers.

Patients' Consents

Given the complexity of ASD diagnoses and that only a subset of patients assessed in the specialized centers are diagnosed as ASD patients, child-psychiatrists at Necker hospital decided to solicit ASD patients' consents when there is an agreement about ASD diagnosis among practitioners in charge of the patient. In this case, further multi-disciplinary investigations are prescribed and some like genetics require specific patient consent.

Patient data access:

TEDIS internet application is accessible only to authorized professionals within identified entities and identified IP address ranges.

A steering committee comprising psychiatric referents in each ASD expert assessment center, regional representative of health ministry, methodologists and physician responsible of TEDIS information system elaborated an ethical chart-agreement to commit all the participants to the TEDIS project.

The chart reminds patient ownership of the data produced in TEDIS as well as the mandatory patients' consents prior to data collection. The chart reminds the contexts of accessing and processing nominative patients' data and anonymized patient data: either for data production and quality control at the production entity site in the first or for aggregated patient data sets for statistical processing in the second case. Besides, the Chart also reminds the role of the scientific committee in selecting research projects and accrediting scientific publications as well as conducting and representing TEDIS project.

We define three contexts for users to access and use TEDIS patient private data:

- Nominative patient data and Data production context: Within an ASD assessment center where explicit nominative patient data, are needed for quality control. Authorized medical and clinical professionals directly participate in TEDIS ASD patients' data production with support of a clinical research assistant (CRA) responsible for quality control and data consolidation. A psychiatric referent designates in each ASD assessment center, professionals au-

thorized to use TEDIS information system and specifies their role. Each authorized professional, may access only patient information s/he is responsible for, within one ASD assessment center. S/He has to sign the chart agreement and is responsible for quality of information produced in TEDIS.

Centralized unique user name and password are issued and maintained by physician responsible of TEDIS at SBIM Necker.

Formal, postal correspondence to authorized professional, communicates access logging keys to use TEDIS, along with chart-agreement for signature and adhesion.

- Anonymized patient data and Data processing context: Only anonymized patient private data are used in this context, for presentations and/or local research projects, within one ASD assessment center or between multiple centers. In this context the CRA will organize periodically meetings with TEDIS steering committee to present the patient data production progress, to comment the data quality control centers and to make suggestions for follow-up.
- Anonymized aggregated patient data may be extracted for broader research projects: Specific scientific and ethical authorizations might be required. TEDIS scientific committee has to decide on how to use TEDIS patient data and specifies conditions to accredit scientific work and publications.

Security infrastructure

Discretionary Access Control

Discretionary Access Control is implemented in TEDIS [4]. Users connect to the application with their user name and password on TEDIS web page [8]. Additional transparent authentication information is controlled when logging into TEDIS system. Number of connection failures is limited and network traffic surveillance permits detecting abnormal behavior and robot phishing [13,14].

We implemented TEDIS based on n-tier architecture and thin internet client. Most of the computing processes occur at the server side. We used Java JSP/Servlets technology and Apache Tomcat Web server to communicate between client and MySQL database server. The servers are deployed on Linux environment [1].

For maintenance and evolution purposes of the security technologies, we referred in our work to standard cryptography methods to reinforce patient privacy protection.

Personal data protection: Scenario of use:

Before submitting the web page form, patient private data are encrypted and sent over secured internet connection to the web server. We used a strong symmetric encoding algorithm (AES with 256 bits keys) in JavaScript, which integrates the time of user's interaction with web-page to generate a single key and cipher text [15] (Figure 1). While building character String to encrypt, we used a standard field separator to distinguish each field (last name, first name, sex and date of birth), and augmented the String with additional information intrinsic to the patient record. The information is sent to the server via secured internet connection.

We decided to strengthen security at the server side, by encrypting again data submitted from the user-interface before storing it on the database server. We first encapsulated each of the cipher text and key respectively into a specific Java Objects called Sealed Object. We used SealedObject Class [9] which enables creating an object and protecting its

confidentiality with a cryptographic algorithm. We then serialized [10,11] the sealed objects and respectively saved them in distributed databases on the server side.

The link between patient private data and clinical data remains the patient ID, automatically generated by TEDIS information System. In case of stolen hard disc with the patients' private data, and to make it more difficult for a non-authorized user to retrace relations between serialized objects and the patient ID, we decided to dynamically generate a one-way hash-coded patient-ID with augmented information and use it to appropriately map serialized objects in the databases [15]. The resulting hash-code-patient-id with the appropriate reference to the serialized objects is saved in separate correspondence database table. Logging access and actions to the correspondence table are particularly monitored within TEDIS application.

Patient data retrieval: scenario of use

To access one patient's clinical record, TEDIS user either enters patient ID or selects patient ID from a list on the web-page. Patient private data retrieval has to go dynamically through the inverse process of encryption, sealed-object creation, and serialization in the database.

A first step is to rebuild the patient-ID-hash-code, then to step back through the distributed databases to select, de-serialize and unseal objects to original encrypted data and original cryptographic key. The latter are sent back to the user interface for decryption with the JavaScript program on the user web page [11]. The patient's corresponding clinical information is directly selected from the database through separate queries and connections and displayed the dynamic web page for display along with the patient private data. The underlying retrieval and decryption processes are transparent. Newly created patient private data are directly encrypted in the database. Basic functionalities of creation, updating and retrieval work in transparent ways with fluidity.

In practice: Encrypting patient private data in TEDIS

Three hundred and sixty ASD patients' records were entered using a prototype TEDIS application on a stand-alone computer server at SBIM Necker. Patients' private data were first saved in plain text form. We updated existing patients' private data using encryption sequence processes described above. Plain text data were kept for verification. Manual control ensured data integrity and correctness.

We were interested in comparing TEDIS performance when saving and retrieving patients' private data in plain text in one database table versus in encrypted serialized objects in distributed databases. We used local copies of the database and programs adapted to the plain text patients' private data for comparisons.

Results

Security policy

Hosting TEDIS within the APHP infrastructure as an official Accredited Hosting Provider in Health Personal Data is a major issue for durability and reliability of our project in terms of security, maintenance, and institutional backing.

TEDIS Chart agreement was elaborated with the steering committee. It represents an ethical and organizational commitment support for all TEDIS users. It had large approval among medical and clinical communities involved in the project, regional health ministry representative and regional

Association for ASD assessment centers where autistics patients' families are represented.

TEDIS in the local stand-alone computer server environment showed correct, consistent, and fluid functioning of the ASD patient private data encrypting-decrypting process for creating a new patient, updating or retrieving patient private data. Comparison (Table 1) showed expected results but gives more precision, an important amplification when scaling (first line 5 seconds versus 81 milliseconds). Retrieving a complete patient record result is not expected, but we are in the time-range of milliseconds, while updating one patient record takes longer with encrypted data.

Table 1 – Performance comparison between one table saved plain text patient private data and distributed database tables saved encrypted patient' private data

Function	Plain text	Encrypted
Retrieve-display a list of 360 records	81 milliseecs	5354 milliseecs
Retrieve-Display one patient record (private data and clinical information)	85 milliseecs	55 milliseecs
Update one patient private data	0 milliseecs (69355nanosecs)	31 milliseecs (31476803nanosecs)



Figure 1 – The presentation layer (4), the style layer (3), the action layer (2) of the html web page layer (1). Information is unlocked in plain text at the presentation layer. The encryption algorithm dynamically integrates the interaction time. The resulting encrypted information is submitted to the server via secure Internet connection.

Discussion

While security policy fits with largely described approaches, we would like to discuss the private data encryption impact on the server database for TEDIS use in clinical setting environment.

In our model we processed multiple encryption processes to offer higher patient private data security, integrity and fluidity and efficiency of the solution.

A first consequence of encrypting patient private data on the server side, is to limit patient data phonetic searches only on the application user interface. To access patient record, users either have to enter known patient ID number or select it from a list of decrypted identities of patients assessed in one assessment entity.

While updating and retrieving one patient record remains in range of milliseconds, displaying a list of more than 300 patient private data after decryption processes, ranges between few seconds. It is still acceptable but borderline display waiting time. The main reason for slowness when scaling is multiple databases connections that have to be processed sequentially and properly.

The client desktop computer also needs to perform to permit fluent encryption-decryption. The JavaScript decryption process is still running sequentially on the web-page while clinical information for all patients is already there ready for appropriate display. Program and database optimisation are needed to improve performance when scaling.

Reducing number of encryption process might be another choice but it will weaken privacy security on the server side. We ended up with the encrypting solution presented in the article, as it showed consistency and stability compared to less elaborated solutions where we either had corrupted data on the server-side or unstable interface decryption.

Conclusion

The security policy, together with the security infrastructure provides a robust and reliable framework for a system like TEDIS to be used in clinical setting, and accessed via secured internet connection.

Acknowledgements

Medical experts and health professionals at the nine ASD assessment centres in Ile De France district and the Autism Resource Centre in Ile-de-France are warmly thanked for their support and feedback. This work is supported by the University of Paris Descartes, The Regional Health Agency of Ile-de-France and by the Necker Hospital and the Computing Department of Assistance Publique des Hopitaux de Paris.

References

- [1] M. Ben Said, L. Robel, E. Vion, B. Golse, J.P. Jais, and P. Landais, TEDIS: an information system dedicated to patients with pervasive developmental disorders, *Stud Health Technol Inform* **160**(Pt 1) (2010), 198-202.
- [2] M. Ben Said, L. Robel, C. Messian, Y. Craus, J.P. Jais, B. Golse, and P. Landais, Patient information, consents and privacy protection scheme for an information system dedicated to pervasive developmental disorders. *Stud Health Technol Inform* **205** (2014), 755-759.
- [3] <http://www.dsm5.org> last consulted on Dec 9thst 2016.
- [4] W.A. Al-Hamdani, Cryptography based access control in healthcare web systems, in proceedings of Information Security Curriculum Development Conference, InfoSecCD'10, ACM New York, NY, USA (2010), 66-79.
- [5] Décret n° 2006-6 du 4 janvier 2006 relatif à l'hébergement de données de santé à caractère personnel et modifiant le code de la santé publique (dispositions réglementaires). JORF n°4, 5 Jan 2006, 174, texte n°14.
- [6] S.B. Johnson, G. Withney, M. Mc Auliffe, H. Wang, E. Mc Creedy, L. Rozenblit, and C.C. Evans, Using global unique identifiers to link autism collections, *JAMA* **17**(6) 2010, 689-695. doi: 10.1136/jamia.2009.002063.
- [7] R.R. Schell, Privacy and security, cyber defense triad for where security matters, *Communications of the ACM*, **59**(11) (2016), 20-23.
- [8] M. Jayabalan and T. O'Daniel, Access control and privilege management in electronic health record: a systematic literature review, *J. Med Syst* **40**(12) (2016), 261.
- [9] M. Ben Said, L. Le Mignot, C. Mugnier, J.B. Richard, C. Le Bihan-Benjamin, J.P. Jais, D. Guillon, A. Simonet, M. Simonet, and P. Landais, A multi-source information system via the internet for end-stage renal disease: scalability and data quality, *Stud Health Technol Inform*, **116** (2005), 994-999.
- [10] P. Landais, C. Messiaen, A. Rath, L. Le Mignot, E. Dufour, M. Ben Said, J.P. Jais, L. Toubiana, G. Baujat, E. Bourdon-Lanoy, M. Gérard-Blanluet, C. Bodemer, R. Salomon, S. Aymé, M. Le Merrer, and A. Verloes, CEMARA: an information system for rare diseases, *Stud Health Technol Inform* **160**(Pt 1) (2010), 481-485.
- [11] N. Tyagi and A. Ganpati, Comparative analysis of symmetric key encryption algorithms, *Int J Adv Res Comput. Sci Softw Eng*, **4**(8) (2014), 348-354.
- [12] A. Czeskis, M. Dietz, T. Kohno, D. Wallach, and D. Balfanz, Strengthening user authentication through opportunistic cryptographic identity assertions. In proceedings ACM – conference on Computer and Communication security - CCS'12, Raleigh, North Carolina, USA (2012), 404-414.
- [13] C. Lu, Z. Wu, M. Liu, W. Chen, and J. Guo, A patient privacy protection scheme for medical information systems, *J Med Syst* **37**(6) (2013), 9982.
- [14] <https://www.fourmilab.ch> last consulted on Dec 9th, 2016.
- [15] <https://docs.oracle.com/javase/7/docs/api/java/lang/Object.html> last consulted Dec 9th, 2016.
- [16] Available at <https://docs.oracle.com/javase/7/docs/api/javax/crypto/SealedObject.html> last consulted on Dec 9th, 2016.
- [17] <https://docs.oracle.com/javase/tutorial/jndi/objects/serial.html> last consulted on Dec 9th, 2016.
- [18] Available at <https://docs.oracle.com/javase/8/docs/technotes/guides/serialization/index.html> last consulted on Dec 9th, 2016.

Address for correspondence

Dr. M. Ben Said, Service de Biostatistiques et d'Informatique Médicale, Hôpital Necker Enfants Malades, 149 rue de Sèvres, 75015 Paris, France
e-mail : mohamed.ben-said@ext.parisdescartes.fr

Mapping the Electronic Health Record: A Method to Study Display Fragmentation

Yalini Senathirajah^a, Jinglu Wang^b, Elizabeth Borycki^c, Andrek Kushniruk^c

^a Department of Medical Informatics, SUNY Downstate Medical Center, Brooklyn, NY, USA,

^b Weill Cornell Medical School, Cornell University, New York, NY, USA

^c University of Victoria, Victoria, Canada

Abstract

Electronic health records have often been criticized for poor interaction design. One major problem is the 'display fragmentation problem' i.e. the fact that conventional EHRs require the user to view many screens and retain information in memory or external tools rather than being able to view all relevant information together, increasing cognitive load and the possibility of errors and inefficiency. We describe a method for evaluating and depicting the extent of display fragmentation and discuss its potential uses in comparing systems, identifying navigation pathways and information juxtaposition, and improving EHR interaction design.

Keywords:

Electronic Health Records; User-Computer Interface; Computer Graphics.

Introduction

Healthcare information technology ('health IT') and electronic health records (EHRs) have great promise to improve care, reduce costs, and create a 'learning healthcare system' in which continuous improvement is possible by using data to analyze which treatments are most effective. However, optimal interaction design of such software has proven difficult, with the potential for health IT to introduce safety concerns.

The Institute of Medicine 2011 report [1] identifies several concerns related to fragmented displays and the conventional interaction approach in which information location is fixed by the programmer and users navigate through menus. These concerns include the mismatch between programmer assumptions and actual work environment and the mismatch between developer and clinician backgrounds, resulting in unmet needs. Current displays may not reflect clinical associations, presenting related data separately. Activities are treated as belonging to individual clinicians, instead of to a sociotechnical system with many intercommunicating components and unpredictable ways [1]. Inflexible order sequences may require providers to hold orders in mind while navigating, and time spent on cumbersome data retrieval and remodeling is time taken from other clinical demands [1].

Display fragmentation occurs in conventional systems when the end user must click through and view many different screens or parts of screens in order to view all the relevant clinical information. This necessitates sequential viewing and requires retaining information in memory while other parts of the information are sought. The end user must often then serve as the 'de facto integrating agent' [1], integrating information in mind rather than using external tools. This is a known cause of excessive cognitive load and may impair optimal clinical

reasoning performance and other aspects of clinical cognition. Addressing the display fragmentation problem is therefore an important task in redesigning clinical information systems for better efficiency and cognitive support.

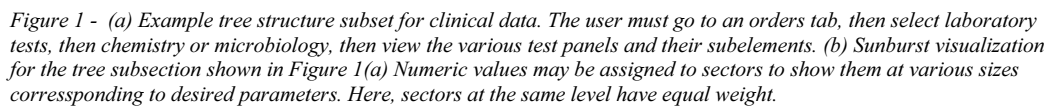
Prior work reveals conventional EHRs have an approximate six-fold greater number of clicks and screen transitions required to obtain complete information [2] than systems in which the relevant information for a task can be included on one screen. Conversely, appropriate information juxtaposition can foster insight, creativity, sensemaking, and problem-solving [3,4].

Modular composable systems

Because the display fragmentation problem arises partly from the fact that information location is typically fixed by the programmer, one means of addressing it is to by making systems a) modular (i.e. pieces can be rearranged and reassembled flexibly according to different needs, like a Lego set), and b) composable by the end user clinician. Removing the requirement for a programmer to reconfigure displays and giving the clinician the ability to assemble any desired elements together on the same screen can allow patient-specific display of all relevant information on the same page, reducing cognitive load and respecting clinicians' deep medical expertise in their choice of elements. Ease of use can mean that drag/drop assembly of 'objects' is a simple means of providing this functionality. Some other domains in which user expertise and security/reliability are important have also used this approach; NASA mission control tools are an example [5].

MedWISER system

Our experimental system, MedWISER, implements a modular composable architecture, in which elements are backed by a controlled vocabulary [6]. Separation of the display elements and back-end data queries allows for a javascript framework which permits the end user to specify display of chosen elements (such as a lab value, note or note fragment, mashup of lab plots, RSS feed) in movable rectangular widgets. These can also be opened to large-screen size, collapsed with only a header showing, colored, retitled, and have other modifications which affect display. This gives the end user considerable power to select and arrange information elements [7]. More advanced features are not discussed in this paper. By permitting choice and assembly of any elements by the clinician, these can be gathered on the same page before or during clinical review of the patient case. This then avoids the fragmentation problem, as repeated navigation to and viewing of other screens is no longer necessary (or minimally necessary). This should avoid the consequent cognitive load, forgetting, navigation inefficiencies and other problems inherent in our current conventional fixed systems. These typically



One way to study and compare systems is to map the locations of the major information elements as they occur according to the navigation structure of the EHRs. This can help us understand the fragmentation problem, suggest design solutions, and compare systems. Here we describe a method of doing this, with a few examples.

Modified Cognitive Walkthrough

Many different visualizations of hierarchical structures exist with advantages and disadvantages. In this technique we have leveraged advantages of the sunburst visualization. This

In the sunburst, sections at the same navigation level appear in the same concentric circle; with leaves below toward the outer rim. Items viewable together on the same screen are included in the same block (with the names all written together in the same sector). Items not viewable together on the same screen

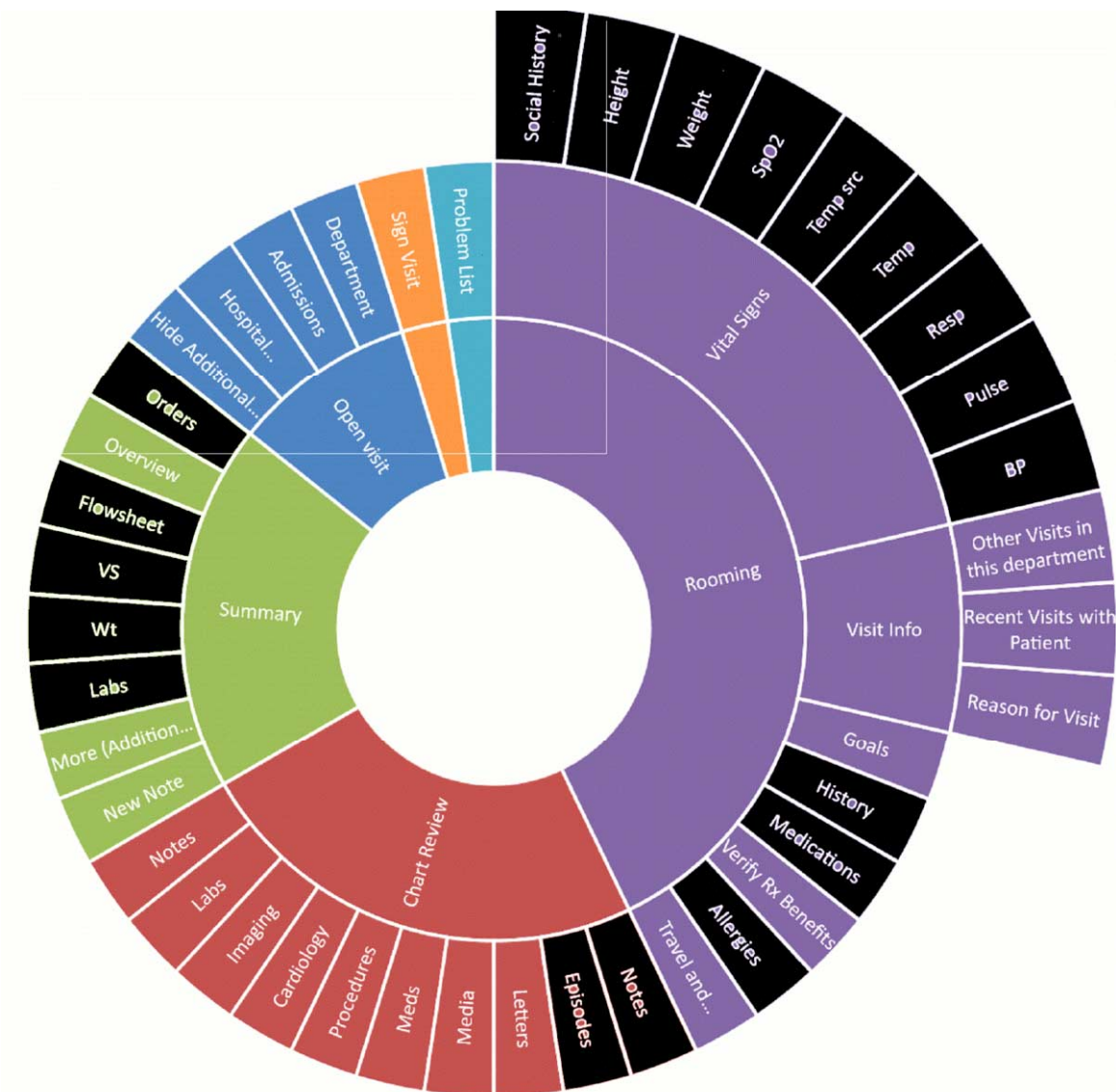


Figure 2 - Sunburst diagram of commercial outpatient system navigation with clinical element leaves shown in black. Each concentric circle represents a level of navigation from the top of the tree. In this case the majority of clinical element values are at least two steps (e.g. clicks) from the origin, and are displayed separately on different screens from other clinical elements, requiring the user to view multiple screens to get a comprehensive clinical overview.

but only sequentially as parts of separate screens accessed from the same menu level appear as separate sectors adjacent to each other, in the same colour as the higher menu level.

Figures 2 and 3 show the essential clinical elements coloured in black. Figure 3 shows the extent of display fragmentation; with each element in a separate sector if it must be navigated to separately, permitting analysis of the EHR structure and where elements are juxtaposed or not.

While figures shown here are relatively simple, visualization software is capable of rendering hierarchies with thousands of elements, with drill-down and expansion capabilities to facilitate examination of relationships between elements, pathway identification, and the study of subsections.

Quantification of Display Fragmentation

The fact that clinical elements appear in different subsections of the tree can be quantified, with respect to useful parameters such as the number of clicks or screen transitions required to access individual elements (starting from the top of the tree, or from intermediate points), as well as the likely numbers of clicks/screen transitions required for typical case review.

Derivation of an expression incorporating these parameters (including such things as requirements to scroll, or filtering functions) could allow comparison across commercial systems, and comparison of whether an interaction design improvement has successfully reduced the numbers of actions and fragmentation.

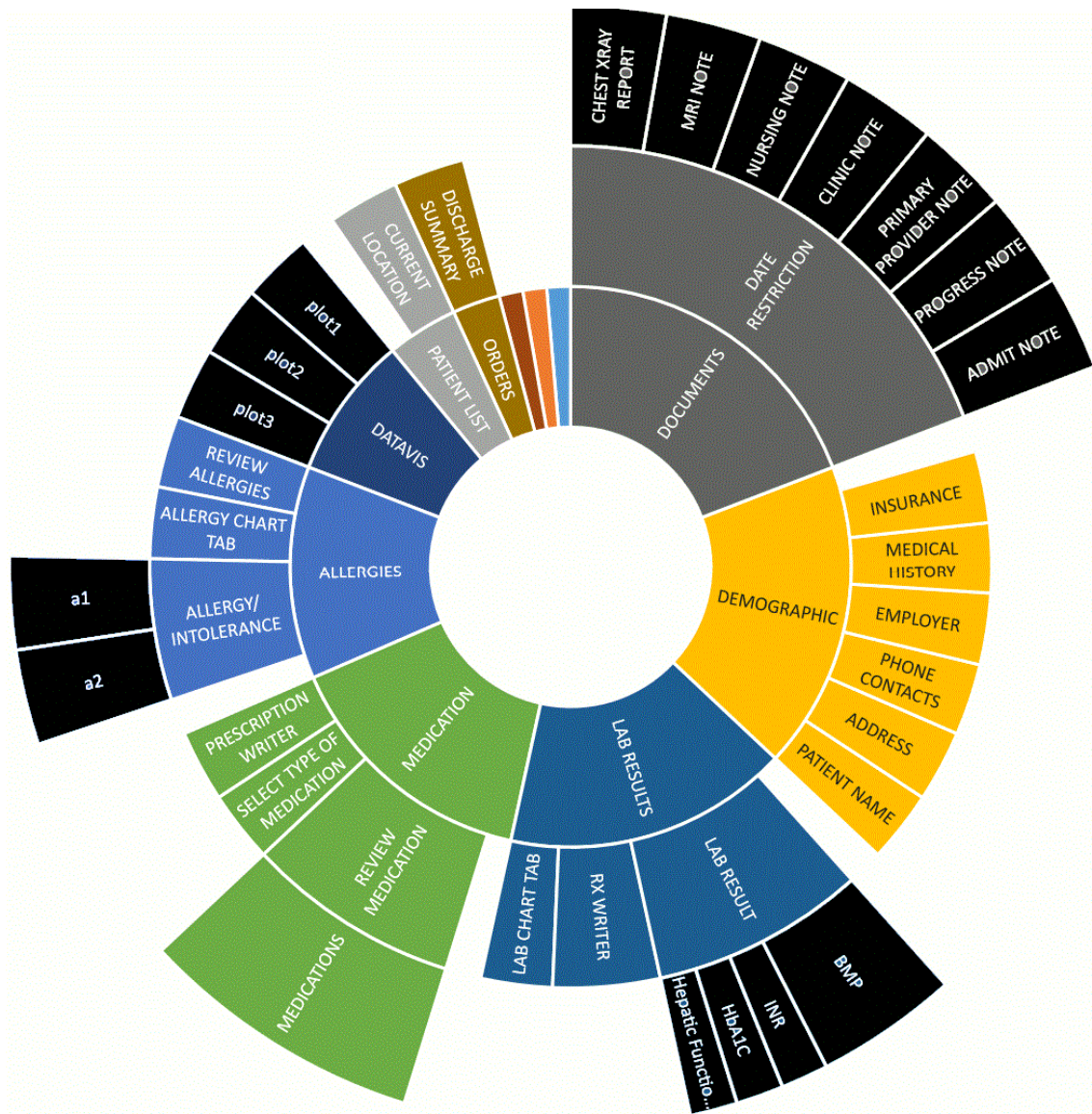


Figure 3 - Display fragmentation in a commercial inpatient system. (Clinical review elements only).
A total of 37 clicks is required to access all the elements shown in black.

Results

Initial trials of this method revealed that a majority of end clinical elements (end leaves) of the navigation tree are located at least two levels below the top, and that consequently numbers of clicks and screens may reach dozens in order to view all relevant information. This echoes complaints by clinician organizations that ‘too many clicks’ are plaguing physicians in practice and delaying rather than hastening good use of the EHR and expected efficiencies [9,10].

Some systems attempt to reduce the problem by having pages for summaries, with varying degrees of success. Two problems identified in our preliminary work are the lack of inclusion of important clinical information in the summary, and the need for scrolling within elements in the summary,

with uniform presentation of blocks of data regardless of whether the blocks are populated or not. This has the potential of wasting space and confusing the user, who may be unaware that the data exists elsewhere. The display fragmentation problem still exists where the user must take an action to view all the data on screen. Roman et al. describe the significance of within-page and between-page navigation as an important factor for EHR usability [11]. An example of the first problem is an outpatient system in which the clinical summary page includes demographic, billing, encounter and allergy data, but not other important clinical variables.

Discussion

Comparative studies across commercial and home-grown EHRs are under way, in order to identify the extent of fragmentation as a problem in general, and specific design

patterns which may contribute to it or conversely reduce fragmentation for specific tasks. The advent of mobile EHRs has also affected the fragmentation in both positive and negative ways. As mobile screens typically display less information, designers may have paid more attention to screen flow in order to ensure correct distribution of clinical values in a way meaningful and convenient to physicians. They are also able to make use of design patterns such as sequential tree menus to focus actions on immediate tasks without extraneous data. On the other hand, if poorly done, this can increase fragmentation, given the screen size limits and need to cater to finger- rather than mouse-driven interaction, which typically has smaller resolution.

The modified cognitive walkthrough method combined with sunburst display has several advantages for data collection. Cognitive walkthrough for the purpose of defining navigation structures is easy to teach, and easily grasped by non-researchers, allowing for others such as clinicians with access to systems to be data collectors. This can facilitate broader studies as specialized access need not be arranged. The use of Excel or similar spreadsheets is familiar to most people, and the branching structure, once explained, is also easy to understand. Automatic generation of sunburst as one of the now standard visualizations allows for easy experimentation. Formatting options allow manual coloring of subsectors. Some visualization tools generate the path from the top of the tree as a part of the legend, changing interactively as the user moves over different elements.

In preliminary studies, users are found to be mostly unfamiliar with the sunburst (unless their particular work requires it), but find it relatively simple to understand once its tree structure basis is explained. Prior empirical work on comparative information visualization of hierarchical structures has shown sunburst visualizations to be more efficient for tasks involving perception of the hierarchy [12], with a shorter learning curve.

Interaction Design Implications and Other Use Cases

Mapping can support composable approaches' reduction of fragmentation, and assist creation of safer design patterns, such as replacements for extensive dropdowns or other undesirable interaction features. Radial menus, for example, leverage humans' good capacity to distinguish small angles. Mapping of redesigns can provide a measure of progress. Multiple workflow paths could be overlaid on the map, as a way of specifying efficient transitions.

EHR display fragmentation maps have other uses besides research. Interactive sunburst visualizations can be used to convey the navigation paths to particular elements, facilitating instruction. It is conceivable that use of maps for comparison of different EHRs might be used for purchasing processes, as part of evaluation of whether the system flow can be made to match workflow needs. In some systems, changes do not propagate to all areas automatically, and a map could facilitate any manual changes necessary. Development uses include to give initial overviews to developers or researchers seeking to improve workflow, and assessing the effect of software modifications on navigation structures, and thus likely effects on time, efficiency, and fit to task.

Conclusions

Display fragmentation is a critical problem for the usability of the electronic health record, and mapping EHR navigational structure with visualization presents useful ways to understand and address it. In addition such maps can be useful for related purposes such as instruction and software development. The

method will allow comparison of fragmentation across EHRs, facilitating our understanding.

Acknowledgements

This work is funded by AHRQ 1R01HS023708-01A2.

References

- [1] Committee on Patient Safety and Health IT, *Technology, Health IT and Patient Safety: Building Safer Systems for Better Care*, Institute of Medicine, 2011.
- [2] Y. Senathirajah and S. Bakken, Visual clustering analysis of CIS logs to inform creation of a user-configurable Web CIS interface, *Methods Inf Med* **50** (2011):337-348.
- [3] A. Kerne, E. Koh, Representing Collections as compositions to support distributed creative cognition and situated creative learning, *New Rev Hypermedia M* **13** (2007):135-162.
- [4] A. Kerne, E. Koh, S. Smith, H. Choi, R. Graeber, A. Webb, Promoting emergence in information discovery by representing collections with composition, In: *ACM SIGCHI conference on Creativity and Cognition Proceedings*, 2007.
- [5] J. Trimble, T. Dayton, and M. Quinol, Putting the users in charge: A collaborative, user composable interface for NASA's Mission Control, In: *Hawaii International Conference on System Sciences*, 2012.
- [6] Y. Senathirajah and S. Bakken, Architectural and usability considerations in the development of a Web 2.0-based EHR, *Stud Health Technol Inform* **143** (2009):315-321.
- [7] Y. Senathirajah, D. Kaufman, and S. Bakken, The clinician in the Driver's Seat: Part 2: Intelligent Uses of Space in a User-composable electronic health record, *J Biomed Inform* **52** (2014):177-188.
- [8] A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, *J Biomed Inform* **37** (2004): 56-76.
- [9] TooManyClicks.com, accessed December 2016.
- [10] J.S. Ash, C.M. Kilo, M. Shapiro, J. Wasserman, C. McMullen, and W. Hersh, *Roadmap for Provision of Safer Healthcare Information Systems: Preventing E-latrogenesis*, Institute of Medicine, Washington, DC, 2010.
- [11] L.C. Roman, J.S. Ancker, S.B. Johnson, Y. Senathirajah, Navigation in the electronic health record: A review of the safety and usability literature, *J Biomed Inform* **67** (2017):69-79.
- [12] J. Stasko, R. Catrambone, M. Guzdial, K. McDonald, An evaluation of space-filling information visualizations for depicting hierarchical structures, *Int J Hum-Comp St* **53** (2000):663-694.

Address for correspondence

Corresponding author: Yalini Senathirajah yalini@zoho.com

Strengthen Cancer Surveillance in Sri Lanka by Implementing Cancer Registry Informatics to Enhance Cancer Registry Data Accuracy, Completeness, and Timeliness

Kamal Seneviratne^a, Sudath Samaraweera^a, Eshani Fernando^a, Suraj Perera^a, Asanga Hemarathna^a

^a National Cancer Control Programme, Ministry of Health, Colombo, Sri Lanka

Abstract

Cancer surveillance is a process of systematic, continuous collection, storage, analysis, interpretation and dissemination of epidemiological information on cancer cases occurring in a particular geographic area. The process of cancer surveillance is also called 'cancer registration'. A system to conduct cancer registration is called 'cancer registry'. The purpose of the project is to implement cancer registry informatics to enhance electronic cancer registry reporting activities that take advantage of emerging health information technology. The expected function of the application is to strengthen the cancer registry data accuracy, completeness, and timeliness. The goals to be achieved are: determine the incidence of cancers with respect to geographic, demographic characteristics; monitor trends and patterns of cancer incidences over time; identify high-risk populations; provide data for epidemiological studies and prioritize health resource allocations.

Keywords:

Registries, Data Collection, Health Information Systems

Introduction

National Cancer Control Programme (NCCP) of Sri Lanka was established in 1980 based on the recommendations made by World Health Organization. NCCP is the national focal point for prevention and control of cancers in the country. The institution is responsible for policy, advocacy, monitoring and evaluation of prevention and control of cancers, conducting surveillance of cancers and facilitating research related to cancer.

Cancer Burden in Sri Lanka

A total of 13,635 new cancer cases had been diagnosed in 2007 [1]. Out of 13,635 new cases, 6,356 were males and 7,279 were females. In 2007 the overall crude cancer incidence rate (CR) was 68.0 per 100,000 population. There were 16,888 new cases diagnosed in 2009 with a CR of 82.6 [2]. In 1985 there were 5,012 new cases with a crude cancer incidence rate (CR) of 31.6 [1]. It is obvious that cancer prevalence is gradually increasing in Sri Lanka. This increase may be due to several reasons [3]. Since a considerable number of government and private health care institutions have started diagnosis and treatment of cancer patients, collection of data from these institutions and statistics related to cancer deaths from death registrars have become absolutely essential to include all these data to fulfill the four quality indicators recommended by the International Agency for Research on Cancer (IARC). They are comparability, completeness, validity and timeliness of information [4].

Importance of Cancer Surveillance

Cancer surveillance provides a quantitative portrait of cancer and its determinants in a defined population. It helps epidemiologists in descriptive studies and analytical studies on cancer [5]. Knowledge of trends in the incidence of cancer over geographical areas may be used for the projection of future incidence rates, case loads, and the need for treatment facilities. It contributes to the management of cancer care programmes to ensure that all patients with a given cancer are given the state-of-the-art diagnosis and treatment.

In cancer surveillance, cancer registries are definitive and unique resources for measuring the cancer burden in a community.

Cancer Registries and Cancer Registration

Cancer registry is a fundamental tool to controlling cancers in a given country. The cancer registry is the office or institution which attempts to collect, store, analyze and interpret data on persons with cancer and cancer registration is defined as the process of continuing systematic collection of data on the occurrences and characteristics of cancer with the purpose of helping assess and control malignancies in the community [6,7]. Cancer registries are an essential part of a complete cancer control program and help in finding etiological factors of cancers primary, secondary and tertiary prevention and monitoring [8,9].

There are three types of cancer registries available globally based on characteristics, purpose, and utility in cancer control. They are:

- a) Hospital-based cancer registry (HBCR) [8,9]
- b) Pathology-based Cancer Registry [8,9]
- c) Population-based cancer registry (PBCR). Cancer information is collected systematically on all reportable neoplasms occurring in a geographically-defined population from multiple sources. These registries have a unique role in planning and evaluating cancer control programs [8,9].

Cancer Surveillance Informatics

The definition of public health informatics is given as the systematic application of information and computer science and technology to public health practice, research, and learning. It is a subdomain of the biomedical or health informatics [10].

Likewise, it is justifiable to use the same terminology in the cancer surveillance informatics domain. Therefore, it deals with the resources, devices, and methods to optimize the acquisition, storage, retrieval and use of information in cancer surveillance practices, research and learning [11].

The tools used in the domain include computers, information and communication systems, clinical guidelines, and formal medical terminologies [11]. Thus, using emerging technology to incorporate automated process and electronic data exchange

in cancer surveillance business is an efficient, fast and cost-effective way to obtain quality, accurate and complete cancer registry data as it increases data accuracy, completeness, timeliness, and comparability. It also minimizes manual handling of documents and manual processing of information.

The National Cancer Control Programme (NCCP) has been monitoring trends of cancer since 1985 and has been publishing the national cancer incidence data. Presently, the National Cancer Control Programme collects data (Figure 1) from case-finding sources including,

- Nine provincial cancer treatment centers. These include Oncology units (9) and oncological surgery units (9). These provincial cancer treatment centers are located in 9 provinces.
- Sixty-seven Histopathology and Hematology laboratories.
- Twenty-five Oral and maxillofacial units.
- Fifteen Government hospitals with specialist care, well woman clinics in Colombo district.
- Five private sector hospitals and pathology laboratories located in Colombo district. Received data from 05 sources.
- 42 death registrars' offices located in Colombo district.

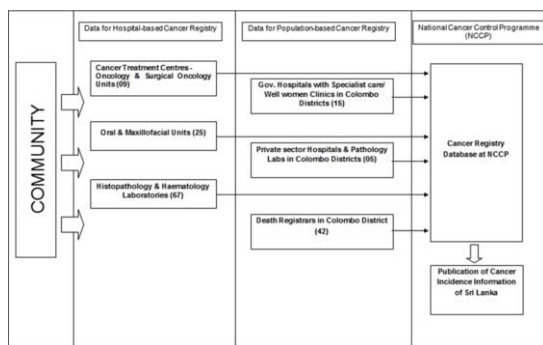


Figure 1. Current cancer registry data flow

Therefore, intended users of the database are the same institutions mentioned above. The current process of cancer registry data gathering from provincial cancer treatment centers as stated above is through an electronic application that was developed by NCCP and installed in a stand-alone desktop computer in each provincial cancer treatment center. The user export data as an excel spreadsheet, which is then sent to the NCCP quarterly. Data received from reminder case-finding sources are based on both paper-based and excel spreadsheets. Data received are then fed into another open source application called CanReg5 [12] for consistency check.

However, with the existing data collection method, the gap between diagnosis of cancer and the availability of data for analysis is significant. As a result, the publication of cancer registry data may reflect information about patients diagnosed more than a couple of years ago. In order to evaluate and react more effectively to trends, there is the need for availability of data on time. Data redundancy is also reported as the same patient attends two clinics in different provinces in the same month. It directly affects the quality, and reliability of data. The issues that are encountered in data gathering process have to be rectified through communication and information technology. Using information technology (IT) in order to shorten all phases of cancer registration process is well documented [13]. Thus, using emerging technology to incorporate automated process and electronic data exchange in cancer surveillance business is an efficient, fast and cost-effective way to obtain quality,

accurate and complete cancer registry data as it increases data accuracy, completeness, timeliness, and comparability.

Methods

A web-based cancer registry information management system called Sri Lanka Cancer Registry (SLCR) was designed, developed, and implemented to maintain all reportable cancer cases using web 2.0 technology, open-source software, and mobile technology. Open-source license products were used to decrease the cost for the technical platform and also as a defining factor in terms of profitability. Although there is no clear-cut definition, the term web 2.0 suggests an upgraded and updated version of the web [14]. Web 2.0 is associated with web development and design. It facilitates interactive information sharing, user-centered design, and collaboration on the World Wide Web[13]. The web 2.0 is able to integrate different sources for a real-time cancer registry. The java technology is used as the tool for the web development process. Java is an object-oriented server-sided programming language and is more suitable for both web and desktop application development and is enabled to write more secure programs [15]. The server is a Linux-based operating system with the Apache as the webserver. The database system for storing information is MySQL, which is multithreaded and multi-user database management system with more than 10 million installations [16]. It is distributed under a GNU General Public License (GPL) and hence has no running cost. The community involvement makes the software development progress to be steady [17]. MySQL can run over 20 operating systems including Linux-based and Windows. Its performance, scalability, and security system have made the system to be used by many large companies world-wide [17].

The software architecture of the Database Management System (DBMS) is client-server architecture. The client-server architecture is considered because this architecture has several advantages (Figure 2).

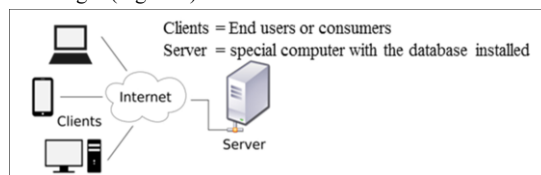


Figure 2. Client-server architecture model

- Centralization:** All the information is stored in a server. Clients can access the database over an authorized access. Every authorized client is given the opportunity to access information via the desktop or laptop interface and tools like spreadsheets. PowerPoint presentations can be used to deal with data with the help of database and application servers resident on the network to produce meaningful information.
- Scalability:** The server where the data are stored can be repaired, replaced, upgraded or relocated while the client or the end users remain unaware and unaffected by that change.
- Availability:** As the server is always turned on, data can be accessed at any time throughout the day and throughout the year.
- Prevent from data losing:** As data is stored on a server it is easy to make a backup. In case of data corruption or data loss, it can be recovered easily and efficiently from backup.
- Security:** Data can be better secured from unauthorized access by defining the permissions at the time of the setup of the server.

There are two models of client-server architecture namely, two-tier and three-tier architecture. From these two options, two-tier architecture is selected for the development of DBMS because it has several advantages over three-tier architecture.

- Two-tier architecture is easy to maintain.
- Modification of the database is relatively easy.
- Communication between the end user and the database is faster.
- Development cost is much less than the three-tier architecture.
- Two-tier architecture is less complex than the three-tier architecture.

The mobile app is used to gather data from death registrars' offices located in Colombo district, Government hospitals with specialist care, well woman clinics in Colombo district, private sector hospitals and pathology laboratories located in Colombo district. As the minimum dataset essential for the cancer registry is collected from these institutions, using of mobile technology is more convenient for them. This method of data collection is greatly supported by the fact that more than 23 million mobile connections exist in 2015 [18]. This figure exceeds the population of Sri Lanka [19].

The process of data collection procedure begins with abstracting data from case-finding sources to a summary sheet, which is designed following several rounds of discussions with stakeholders in order to satisfy their needs yet preserving standard coding and classification procedures and definitions for the population-based cancer registry [21]. It consists of two sections. Section I is related to personal identification and demographic data and is supposed to be completed by the nursing officer of the relevant oncology unit. Furthermore, she can further ascertain data when the patient attends oncology clinic. Section II is related to tumor details and is supposed to be completed by a clinician in relevant oncology unit except ICD-O3 coding. Section II is mostly composed of coded data fields. Just putting the code number in the relevant field is all that is needed. It is expected to improve efficacy on data reporting by releasing extra burden for busy oncology clinicians engaged in registry activities to some extent. Thus the abstracted data in the summary sheet are then transferred to the electronic system by a Data Entry Operator (DEO) through system's web interface. Figure 3 shows the summary sheet to be used for feeding data to the system.

Figure 3. Summary sheets

Results

The intended results to be expected by implementing cancer registry informatics is to enhance cancer registry data

completeness, accuracy (or validity), and timeliness as well as comparability. A cancer registry of the modern era is a medical registry and fulfill the standards defined by Drolet and Johnson [22]. The standard explained five features. Data from multiple sources are combined to create an aggregate set of data called mergeable data (M). data such collected are standardized data set (D), in that the same set of characteristics is collected for each patient in the dataset according to the defined set of rules (R). Furthermore, patients in the registry have a unique identifier in order to follow up over time (O), and this helps to assess the occurrence of specific outcomes of interest (k) [22]. It is expected that the new system will fulfill the criteria of a modern cancer registry.

As the value of any cancer registry relies on the underlying quality of its data and quality control procedures it has in place, the Cancer Incidence in Five Continents (C15) assesses three dimensions of quality to enable the comparison of cancer incidence rate across different population worldwide [23]. They are completeness, comparability, and validity (or accuracy). Completeness is defined as the degree to which all diagnosed neoplasms within a registry's catchment population are included in the registry database [23]. The indices of completeness are grouped into historical data, the proportion of cases microscopically verified (MV%), the mortality-to-incidence (M:I) ratio and death certificate method. Historical data are essential to find out the stability of incidence rates over time, to compare with incidence rates in different populations, to prepare age-specific incidence curves and childhood incidence rates [23]. Therefore the system is designed to include case-finding sources that are currently available with the facility for future expansion. The proportion of cases microscopically verified (MV%) is the percentage of cases that were diagnosed on the basis of microscopic verification of a tissue specimen [23]. This includes histologically, cytologically and haematologically confirmed specimens. As the system is able to integrate those data from relevant laboratories further strengthening the completeness. The cause for death due to cancer in death certificates plays a significant role in the completeness of the registry. It helps to track down the cases not captured by other registration procedures. On the other hand, it also helps to find out Mortality-to-Incidence (M:I) ratio. It is an important indicator of completeness [23] where it compares the number of death due to a specific type of cancer over a specific period of time with the number of new cases of that type of cancer registered during the same period [23]. The M:I ratio that is higher than expected raise the suspicion of completeness in that registry. As the data from death registrars' offices are integrated into the system for PBCR not only enhance the completeness but helps in evaluating the quality of the registry.

Data quality is equal to the completeness of data and its accuracy (or validity)[21]. According to the International Standards Organization (ISO) data quality as "the totality of features and characteristics of an entity that bears on its ability to satisfy stated and imply needs" [23]. In the context of a cancer registry, it is defined as the proportion of cases in a dataset recorded as having a given characteristics that truly has that attribute [23]. Accurate or quality data is essential to compare data between registries within a single registry over time and within a single registry with respect to a specific subset of cases [23]. The proposed web-based system accomplishes data accuracy in several ways. The coding standards for neoplasms are maintained using the third edition of the International Classification for Oncology third edition (ICD-O-3) [24]. It provides standardized system for coding the anatomical site of primary tumor (Topography), and its histological type (Morphology) usually obtained from

pathology report, its behavior (malignant, benign, in situ, or of uncertain behavior), grade (the extent of the differentiation of the tumor) and the basis of the diagnosis (the method of diagnosis used) [23]. Thus the NCCP act as the database administrator and registry staff review each record in the system carefully and verify ICD-O code against topography, morphology, and behavior and also against sex and age. Coding of neoplasms are done only by them and not the duty of the institution where data have been obtained. If data inconsistency found between demographic data and tumor data in a particular record, For example, retinoblastoma is a malignant disease of the retina, exclusively found in young children. If the record shows the morphology as retinoblastoma but the patient's age as 45 years with the tumor behavior as benign or in another record topography has been recorded as prostate carcinoma for a female patient, then the DEO of the respective data center can be notified through the system by NCCP (registry staff). Then DEO of a particular data center is able to communicate with the clinician in order to rectify them accordingly. Figure 4 shows flagging system

Figure 4. Flagging system with example synthetic patient data.

Sri Lanka Cancer Registry (SLCR)

Home

Reporting

Confirmed

Patients

New Patient

Provincial General Hospital Battaramulla (Local Operator)

Clinic File No.	Date of Registration	Surname	Other Names	Date of Birth	Gender	HIC No.	Permanent Address	Contact Address	Status
12345	2016-11-01	SM	Thirabanda	1950-01-01	Male	88908765V	Pennara road, Battaramulla	Pennara road, Battaramulla	Confirmed View/Edit
inccr-08-14	2014-09-07	Fernando	Elan	1970-06-08	Female	Kinla Road, Colombo-05	Kinla Road, Colombo-05	Confirmed View/Edit	
inccr-344-12	2012-12-04	Mariakulasingam	Arunabalan	1939-09-30	Female	Elaka North, Elaka	Elaka North, Elaka	Confirmed View/Edit	
inccr-342-12	2012-10-03	Vinayakumar	Rasmanan	1969-04-18	Female	69607280DV	Nelukulam, Vavuniya	121, Access Road, Battaramulla	Confirmed View/Edit
inccr-22-20-2014	2012-10-25	Kanapathy	Sellamann	1948-12-12	Female	123	Kandy Road, Kandy	Kudalady, Jaffna	Correction View/Edit

Copyright 2016 National Cancer Control Directorate. All Rights Reserved. Version 1.0 (2016)

Figure 5. Confirmed data with example synthetic patient data

Thus, the system assists in maintaining data consistency as well as timeliness as the data can be viewed as soon as they are included in the system and ability to correct them as soon as possible. The stand-alone databases currently used to collect data in case-finding sources are unable to fulfill such advantages. Once confirmed by the NCCP a record as a completed record, it is shown as confirmed record in the system (Figure5: image showing confirmed data).

Discussion

The intended results to be expected by implementing cancer registry informatics is to enhance cancer registry data completeness, accuracy (or validity), timeliness, and comparability. A cancer registry of the modern era is a medical registry and fulfill the standards defined by Drolet and Johnson [22]. The standard explained five features. Data from multiple

sources are combined to create an aggregate set of data called mergeable data (M), data collected are standardized data set (D), in that the same set of characteristics is collected for each patient in the dataset according to the defined set of rules (R). Furthermore, patients in the registry have a unique identifier in order to follow up over time (O), and this helps to assess the occurrence of specific outcomes of interest (k) [22]. It is expected that the new system will fulfill the criteria of a modern cancer registry.

As the value of any cancer registry relies on the underlying quality of its data and quality control procedures, it has in place, the Cancer Incidence in Five Continents (C15), to assess three dimensions of quality to enable the comparison of cancer incidence rate across different population worldwide [23]. They are completeness, comparability, and validity (or accuracy). Completeness is defined as the degree to which all diagnosed neoplasms within a registry's catchment population are included in the registry database [23]. The indices of completeness are grouped into historical data, the proportion of cases microscopically verified (MV%), the mortality-to-incidence (M:I) ratio and death certificate method. Historical data is essential to find out the stability of incidence rates over time, to compare with incidence rates in different populations, to prepare age-specific incidence curves and childhood incidence rates [23]. Therefore the system is designed to include case-finding sources that are currently available with the facility for future expansion. The proportion of cases microscopically verified (MV%) is the percentage of cases that were diagnosed on the basis of microscopic verification of a tissue specimen [23]. This includes histologically, cytologically and haematologically confirmed specimens. As the system is able to integrate those data from relevant laboratories further strengthening the completeness. The cause for death due to cancer in death certificates plays a significant role in the completeness of the registry. It helps to track down the cases not captured by other registration procedures. On the other hand, it also helps to find out Mortality-to-Incidence (M:I) ratio. It is an important indicator of completeness [23] where it compares the number of death due to a specific type of cancer over a specific period of time with the number of new cases of that type of cancer registered during the same period [23]. The M:I ratio that is higher than expected raise the suspicion of completeness in that registry. As the data from death registrars' offices are integrated into the system for PBCR not only enhance the completeness but helps in evaluating the quality of the registry.

Data quality equals to the completeness of data and its accuracy (or validity) [21]. According to the International Standards Organization (ISO) data quality as "the totality of features and characteristics of an entity that bears on its ability to satisfy stated and imply needs"[23]. In the context of a cancer registry, it is defined as the proportion of cases in a dataset recorded as having a given characteristics that truly do have that attribute [23]. Accurate or quality data is essential to compare data between registries within a single registry over time and within a single registry with respect to a specific subset of cases [23]. The proposed web-based system accomplish data accuracy in several ways. The coding standards for neoplasms are maintained using the third edition of the International Classification for Oncology third edition (ICD-O-3) [24]. It provides standardized system for coding the anatomical site of primary tumor (Topography), and its histological type (Morphology) usually obtained from pathology report, its behavior (malignant, benign, in situ, or of uncertain behavior), grade (the extent of the differentiation of the tumor) and the basis of the diagnosis (the method of diagnosis used) [23]. Thus the NCCP act as the database administrator and registry staff

review each record in the system carefully and verify ICD-O code against topography, morphology, and behavior and also against sex and age. Coding of neoplasms are done only by them and not the duty of the institution where data have been obtained. If data inconsistency found between demographic data and tumor data in a particular record, For example, retinoblastoma is a malignant disease of the retina, exclusively found in young children. If the record shows the morphology as retinoblastoma but the patient's age as 45 years with the tumor behavior as benign or in another record topography has been recorded as prostate carcinoma for a female patient, then the DEO of the respective data center can be notified through the system by NCCP (registry staff). Then DEO of a particular data center is able to communicate with the clinician in order to rectify them accordingly. Figure 5 shows flagging system.

Thus, the system assists in maintaining data consistency as well as timeliness as the data can be viewed as soon as they are included in the system and ability to correct them as soon as possible. The stand-alone databases currently used to collect data in case-finding sources are unable to fulfill such advantages. Once confirmed by the NCCP a record as a completed record, it is shown as confirmed record in the system (Figure5: image showing confirmed data).

Conclusion

The web-based application will ensure timely availability of information and it will provide accurate, complete and timely registry data to the stakeholders such as health policy makers, researchers, clinicians, and the general population.

Acknowledgments

International Agency for Research on Cancer (IARC), Health Informatics Society of Sri Lanka (HISSL)

References

- [1] Ministry of Health, Sri Lanka, *Cancer Incidence Data, Sri Lanka 2007*, Colombo: National Cancer Control Programme, Sri Lanka, 2013.
- [2] National Cancer Control Programme, *Cancer Incidence Data Sri Lanka 2015*, Colombo, National Cancer Control Programme, 2015.
- [3] F. Bray, A. Jemal, N. Grey, J. Farlay, D. Forman, Global cancer transitions according to the Human Development Index (2008-2030): a population-based study, *Lancet Oncol* **13** (2012), 790-801.
- [4] Y.A. Leal, L.M. Fernandez-Garote, et al. The importance of registries in Cancer Control, *Salud Publica Mex* **58** (2016), 09-316.
- [5] O.M. Jensen, H.H. Storm, (1991) 'Purposes and uses of cancer registration', in Jensen, O.M., Parkin, D.M., MacLennan, R., Muir, C.S., Skeet, R.G. (ed.) *Cancer Registration: Principles And Methods*. Lyon, France, *International Agency for Research on Cancer*, 13-26.
- [6] S. Rajaraman, An evaluation of Survival of Cancer Patients Based on Registry Data from Low or Medium Resource Countries, Finland, *University of Tampere* (2012).
- [7] O.M. Jason, S. Whelan, Planning a cancer registry. In: *Cancer Registration: Principles and Methods, International Agency for Research on Cancer, Lyon France* (1991), 22-28.
- [8] Y.A. Leal, L.M. Fernandez-Garote, et al. The importance of registries in Cancer Control, *Salud Publica Mex* **58** (2016), 309-316.
- [9] F. Bray, A. Znaor, P. Cueva, et al, Planning and Developing Population-Based Cancer Registration in Low-and-Middle-Income Settings, *International Agency for Research on Cancer, Lyon France*, (2014).
- [10] Centers for Disease Control and Prevention, The Role of Public Health Informatics in Enhancing Public Health Surveillance, *MMWR* **61** (Suppl; July.27.2012), 20-24.
- [11] National Cancer Registrars Association, *Cancer Informatics*, Available at: <http://www.ncra-usa.org/i4a/pages/index.cfm?pageid=3934>, [accessed on: Sep 22 2016].
- [12] International Agency for Research on Cancer (IARC). Available: <http://www.iarc.fr/> (Accessed: 16th September 2015).
- [13] F. Bianconi, V. Bruno, P. Valigi, F. LaRosa, and F. Stracci, Information Technology as Tools for Cancer Registry and Regional Cancer network Integration, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, **42**(6) (2012), 1410-1424.
- [14] D. Bear, Making Friends with Jarvis Cocker: Music Culture in the Context of web 2.0, *Culture Sociology*, **2**(2) (2012), 222-241.
- [15] S. Rabah, J. Li, M. Liu, Y. Lai, Comparative Studies of 10 Programming Languages within 10 Diverse Criteria - a Team 10 COMP6411-S10 Term Report, *Concordia University*.
- [16] Y. Bassil, A Comparative Study on the Performance of the Top DBMS Systems, *Journal of Computer Science & Research*, **1**(1), (2012), 20-31.
- [17] I.S. Stroe, MySQL databases as part of the Online Business, using a platform based on Linux, *Database Systems Journal*, **2**(3/2011) (2011), 3-11.
- [18] Colombo Digital Marketers *Internet Usage Statistics in Sri Lanka*, Available at: <http://www.digitalmarketer.lk/internet-usage-statistics-in-sri-lanka.html> (Accessed: 10 November 2016).
- [19] Department of Census and Statistics- Sri Lanka. Available at: <http://www.statistics.gov.lk/PopHouSat/VitalStatistics/Bulletin/Annual%20bulletin%202015.pdf>, (Accessed: 10 November 2016).
- [20] A. Sulman, Top Ten Database Security Threats, The Most Significant Risks of 2015 and How to Mitigate Them Available at: https://www.im-perva.com/docs/gated/WP_TopTen_Database_Threats.pdf (Accessed: 10th September 2015).
- [21] I.D.S.Silva (1999) *Cancer Epidemiology: Principles and Methods, International Agency for Research on Cancer, Lyon France* (1999), 385-403.
- [22] B.C.Drolet., K.B.Johnson, Categorizing the world of registries. *Journal of Biomedical Informatics*, **41**(2008), 1009-1020
- [23] F. Bray, B. Kohlar, and J. Ferlay, Data Comparability and Quality. *Cancer Incidence in Five Continents Volume X*, **7** (2014), 89-97.
- [24] A. Fritz., C. Percy, et al., *International Classification of Disease for Oncology*, Third Edition, *World Health Organization, Geneva, Switzerland* (2000).
- [25] D. Arts., et al, Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. *Journal of American Medical Informatics Association* **9** (2002), 600-611.
- [26] C.W.Whitney., B.K.Lind., P.W.Wahl, Quality assurance and quality control in longitudinal studies. *Epidemiol Rev*.**20**(1)(1998),71-80.

Address for correspondence

Kamal Seneviratne. Kamalse1@gmail.com

Using Therapeutic Circles to Visualize Guideline-Based Therapeutic Recommendations for Patients with Multiple Chronic Conditions: A Case Study with GO-DSS on Hypertension, Type 2 Diabetes, and Dyslipidemia

Brigitte Séroussi^{a,b}, Alexandre Galopin^c, Mouhssin Gaouar^a, Suzanne Pereira^c, Jacques Bouaud^{d,a}

^a Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, LIMICS, Paris, France ; INSERM, U1142, LIMICS, Paris, France ; Université Paris 13, Sorbonne Paris Cité, LIMICS (UMR_S 1142), Bobigny, France

^b AP-HP, Hôpital Tenon, DSP, Paris, France; APREC, Paris, France

^c Vidal, Issy-les-Moulineaux, France

^d AP-HP, DRCD, Paris, France

Abstract

Clinical decision support systems (CDSSs) have proven to potentially improve the compliance of physician decisions with clinical practice guidelines (CPGs). However, actual patients suffer from multiple conditions and CPGs that are usually single-disease-focused provide disease-specific recommendations with no support on how to manage adverse interactions between the recommended treatments. We have developed GO-DSS, a CDSS that implements an ontological reasoning process to perform CPG reconciliation. GO-DSS is applied to the concurrent management of hypertension, type 2 diabetes, and dyslipidemia. We proposed an innovative graphical interface to display medication recommendations as "therapeutic circles". A qualitative evaluation of the system and of this graphical layout has been performed on simulated patient cases by a sample of 12 users with various backgrounds (think aloud method). The resulting usability of the system is highly appreciated with a mean rating of 90.7% according to the standardized System Usability Scale.

Keywords:

Computer Graphics; Decision Support Systems, Clinical; Decision Making, Computer-Assisted.

Introduction

Clinical practice guidelines (CPGs) are evidence-based recommendations to manage patients with specific conditions (e.g., 2014 Evidence-Based Guidelines for the Management of High Blood Pressure in Adults¹, Standards of Medical Care in Diabetes – 2016², or 2016 ESC/EAS Guidelines for the Management of Dyslipidemia³). Based on the best available research evidence, CPGs are currently developed by health professional societies and national health agencies to improve the quality of clinical care and decrease health care costs. CPGs are usually elaborated as textual narrative documents. They may also include tables and graphs. More recently, the National Institute for Health and Care Excellence (NICE) has started to provide guidelines as structured dynamically interactive pathways⁴.

Despite the wide development and dissemination of CPGs, there are still unwarranted variations in clinical practice [1,2]. Indeed, simply providing CPGs in their original narrative format has proven to have a limited effect in changing physician behavior. Several reviews [3,4] suggest that clinical decision support systems (CDSSs) that provide patient-specific guideline-based recommendations may be efficient tools to promote the adoption of CPGs by physicians. However, although many studies have indeed showed positive effects, others have found only a limited impact of these systems upon physician practices [5]. Delivering patient-specific recommendations at the point of care appears to be "neither necessary nor sufficient" to ensure compliance [6]. Research is thus currently carried out to assess which factors are responsible of the success or the failure of CDSSs [7]. Beyond variations in clinical setting, culture, training, and organization, the aim is to analyze when CDSSs are used and how, in order to elicit the technical features, e.g. design, implementation, level of description, as well as usability and display that would predict their correct use and effectiveness to increase clinician compliance with CPGs. Some authors of this article already studied the patient effect on non-compliance with the ASTI system [8] concluding that for "complex" patient cases, general practitioners (GPs) accept help and on-demand guidance-based systems are more appropriate, whereas for "simple" patient cases, GPs do not think they need to be helped, and automatic alert-based CDSSs are both efficient and mandatory since GPs would not spontaneously seek for information.

However, although the implementation of CPGs in CDSSs may be useful to provide clinicians best patient-centered recommendations to manage a given pathology (e.g. hypertension, type 2 diabetes, or dyslipidemia), this does not solve the problem of improving care quality and overall public health which is the ultimate target of CPG development and dissemination. CPGs are focused on the management of a single disease, whereas multimorbidity is a common phenomenon [9]. This is known for elderly patients but also true for younger patients [10]. Thus providing CDSSs to improve adherence with monomorbidity-focused, mono-disciplinary CPGs for patients with multiple conditions may result in undesirable effects: each guideline provides a recommendation but there is a lack of support as to how to manage adverse interactions between recommended treatments and conflicting management strategies. For people with multimorbidity, current guidelines and recommendations rapidly cumulate to drive polypharmacy without providing guidance on how to compare relative beneficence of risks of treatments according to the severity of the different conditions to best prioritize recom-

¹ <http://jamanetwork.com/journals/jama/fullarticle/1791497>

² http://care.diabetesjournals.org/content/39/Supplement_1/S4

³ <https://academic.oup.com/eurheartj/article/37/39/2999/2414995/2016-ESC-EAS-Guidelines-for-the-Management-of>

⁴ <https://pathways.nice.org.uk>

mentations or to select the recommendations that might be dropped.

To improve the management of patients with multiple conditions, some research is conducted on the *a priori* development of guidelines that account for multimorbidity. Epidemiological strategies based on the provision of a checklist of disease combinations that should be systematically considered during guideline development have been proposed to help guideline developers [11]. Other research teams have proposed to work on the *a posteriori* reconciliation of multiple single-disease CPGs. Wilk et al. [12] have proposed a framework employing first order logic to represent CPGs and to mitigate possible adverse interactions (drug-drug or drug-disease) when concurrently applying multiple CPGs to a multimorbid patient. This mitigation algorithm is used as an alerting tool to support the physician in the concurrent application of CPGs. More recently, this work has been refined to extend the mitigation algorithm and include patient's preferences [13]. Other authors have proposed semantic web ontology-based approaches [14] for the integration of multiple single disease clinical pathways in a unified disease-specific clinical pathway. The execution of the ontological clinical pathway model is achieved through abstraction processes to assign functional behaviors to existing semantic properties and facilitate their execution [15].

Information visualization may be defined as the use of visual representations of data, information, and knowledge to help users gain a deeper understanding of the contents of a domain. Numerous research works are currently being carried out to develop health information visualization techniques expected to increase the benefits of health informatics databases and networks. The underlying principle is that the right display of health information should match the mental constructs and cognitive tasks of the user and thus should reduce the cognitive load of data interpretation. As a consequence, the capacity of patients, clinicians, and public health policy makers to make better decisions should be improved [16]. Indeed, literature shows that different types of graphical information can help or arm the accuracy on decision-making [17]. This has been shown at the population level [18], but also at the patient level with Computerized Physician Order Entry (CPOE). Poorly designed CPOE can lead to usability problems, users' dissatisfaction, and may disrupt the normal flow of clinical activities. Wipfli et al. [19] proposed an alternative strategy to alert display and layout that reduces interruptions to physicians' workflow. Payne et al. [20] recommended using visual cues, minimal text, formatting, content and reporting standards to improve drug-drug interaction alerts. The same conclusions apply to information retrieval of CPGs with an interactive graphical interface using an iconic language [21]. Understanding how to best visualize data, information, and knowledge, especially in CDSSs is a central challenge to improving healthcare.

We have developed GO-DSS, a guideline-based decision support system applied to the management of the cardiovascular risk [22]. GO-DSS uses an ontology-based approach to allow for the flexibility needed to deal with patients with multiple chronic disorders. A first implementation concurrently applying hypertension (HT) and type 2 diabetes (T2D) CPGs was previously developed [23]. The system was then successfully extended to integrate the management of patients with dyslipidemia (D). GO-DSS guideline-based recommendations have been extended to be displayed as "therapeutic circles". This graphical layout of recommended drugs and drug combinations was evaluated by a sample of users with various backgrounds (clinicians, GPs, pharmacists, informaticians, engineers).

Materials and Methods

In this section, we describe the three CPGs we worked with, summarize the main functionalities of the guideline-based GO-DSS CDSS, introduce the therapeutic circles we used in GO-DSS to display recommendations, and present the protocol implemented to evaluate the global system.

Hypertension, type 2 diabetes, and dyslipidemia CPGs

We used CPGs synthesized by Vidal⁵, a French company that markets a drug database and medical situations which have been evaluated for quality by medical professionals for decades. CPGs were manually translated into IF-THEN decision rules. THEN-parts contained "possible" actions, "recommended" actions, and "contra-indicated" actions according to CPGs. The 2016 dyslipidemia CPG was translated following the same method and format as the HT and T2D CPGs [22]. We therefore had generated three rule bases, for HT, T2D, and D CPGs.

GO-DSS ontological reasoning

The first step was to build an ontology of the cardiovascular domain. We reused an emergency care ontology (Onto-IUrgences⁶) where concepts relevant to cardiovascular risk management were extracted. This first ontology was then enriched to integrate the CPG-specific concepts used by the decision rules (IF- and THEN-parts) of the three rule bases. Concepts were structured by subsumption, equivalence, and disjunction relationships.

IF-parts of rules are logical expressions (mainly conjunctions) built with concepts from the ontology. The set of IF-parts of rules represent the set of theoretical patient profiles covered by CPGs. Each theoretical patient profile is equivalent to a new concept in the ontology, and as such can be classified by the ontological reasoner. Therefore, these profiles, and their corresponding rules, are organized in a subsumption graph, with the least specific profiles at the top and the most specific ones at the bottom.

At execution time, when an actual patient's case is considered, patient data is encoded as a conjunction of concepts of the ontology to build the closest formalized patient profile. Processing the ontological reasoning consists of identifying all the CPG-based, rule-issued, patient profiles that subsume the formalized patient profile to collect the decision rules that apply to the patient. All the linkages between co-illnesses covered by CPGs are returned by the ontology. However, knowledge gaps in CPGs still produce missingness errors in the reasoning process. The subsumption graph of patient profiles/rules is also used to solve potential conflicts between inferred actions (within or across CPGs) by selecting the actions recommended by the most specific rules in the subsumption graph of profiles and by eliminating the actions of the more general rules, thus implementing a kind of non-monotonic reasoning [22].

As a result, we get all the actions (either "possible", "recommended", or "contra-indicated") filtered by the conflict resolution process that apply, for the best management of the patient.

Display of recommendations

Therapeutic recommendations that are issued from CPGs for a given patient correspond to possible, recommended, or contra-indicated drug classes and drug combinations. In order to summarize all these drug-related recommendations for a given

⁵ <https://www.vidal.fr/recommandations/>

⁶ <https://bioportal.bioontology.org/ontologies/ONTOLURGENCES>

pathology/CPGs, we have introduced a graphical representation we called "therapeutic circles". A therapeutic circle corresponds to the circular disposal of discs where each disc represents one of the drug classes that can be prescribed for a given pathology (HT, T2D, and D) according to CPGs. When the drug class is just mentioned in CPGs, with no other inferred information, the disc is colored in grey. The disc is colored in yellow when the drug class has been mentioned as "possible" for the patient, in green when it is "recommended", and in red when it is "contra-indicated". When the disc has a bold outline, this indicates that the corresponding drug class is currently administered to the patient. Lines between therapeutic discs describe the combination of drug classes with the same color rules. Yellow lines indicate the combination is "possible", green lines indicate the combination is "recommended", and red lines indicate the combination is "contra-indicated". Figure 1 displays the therapeutic circle to represent guideline-based therapeutic propositions for a patient currently treated by ACE inhibitors (bold circle) for whom it is recommended to go for a bitherapy by adding a thiazide diuretic (green link). Adding calcium channel blockers (CCB) or a beta-blocker is possible (yellow links), but adding an angiotensin II receptor antagonist (ARB) is contra-indicated (red link).

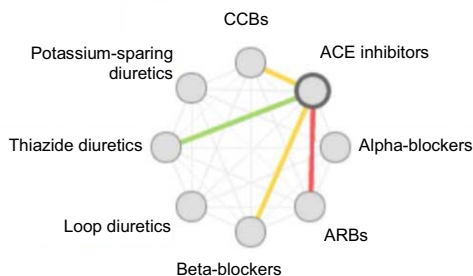


Figure 1 – Antihypertensive drug recommendations represented as a therapeutic circle

For a given patient, multiple pathologies/CPGs must be considered simultaneously. Following the principle of having patient problem lists in medical records as a way to "encourages doctors to think holistically about their patients" [24], we have structured the display of guideline-based decision support propositions per problem, with one column for HT, one for T2D and one for D. Within each column, we deliver six levels of information:

1. The description of the current treatment with the combination level and the list of drugs and their class,
2. The assessment of the current treatment in terms of compliance with guidelines and therapeutic efficacy,
3. The recommended level of drug combination (no drug, mono-, bi-, tri- and quadritherapy),
4. The therapeutic circle to represent guideline-based drug recommendations,
5. Indications about alerts, risks, and surveillance,
6. Non-pharmacological recommendations.

Figure 2 displays the six levels of information for a man of 84 years with asthma and uncontrolled HT despite a bitherapy of antihypertensive drugs, ACEi and a thiazide diuretic, a non-efficient metformin monotherapy for T2D, and a dyslipidemia

non-currently treated. The current treatment globally complies with CPGs but is not efficient:

- For hypertension, a tritherapy is recommended, but a quadritherapy is contra-indicated. The recommended tritherapy is made of the combination of ACEi, thiazides diuretic, and calcium channel blocker (CCB). ARBs are recommended as a therapeutic drug class (green disc) but not in association with ACEi (red line). Beta-blockers are contra-indicated because of asthma (red disc) as well as any combination including them (red lines).
- For type 2 diabetes, a bitherapy made of a combination of metformin and a sulphonylurea is recommended.
- For dyslipidemia, a monotherapy is recommended, by either statins or ezetimibe.

Evaluation protocol

A pilot evaluation of the system has been made off-line for test cases. The evaluation protocol comprised four steps: (i) a tutorial to introduce GO-DSS, (ii) a training step, (iii) the unsupervised use of the system on simulated patient cases, and (iv) a qualitative user assessment through questionnaires. A whole evaluation session was expected to last about 45 minutes.

The first step presented the aim of the system, the context of its development, its main functionalities, and how to recognize and interpret the different types of information displayed in the user's interface (UI). During the second step, the user utilized the system under the supervision of the evaluator on commented simulated patient cases. The aim was to make the user explore and discover GO-DSS functionalities. During this step, the user could ask any question about the system. The third step consisted in the autonomous use of the system by the user alone, without supervision. Two simulated patient cases were proposed. These cases corresponded to the longitudinal management of patients and included multiple consultations, the follow up of evolving chronic conditions, and multiple decision points with drug prescription and adaptation. Such cases were built to illustrate the different operating functionalities of the system. User prescriptions were collected, as well as any enunciated remarks (think aloud method). The user could exit the scenario planned for the simulated cases and he/she was permitted to try any arbitrary patient conditions to test the system response. At the fourth and last step, the user filled a questionnaire to indicate his/her professional profile, as well as his/her familiarity with CPGs and decision support tools. Questions about the medical relevance of recommendations, the consistency with the original CPGs, and the perceived utility were encoded using a 4-valued Likert scale. It was also possible to enter further comments about the system. Finally, a standardized SUS (System Usability Scale) [25] questionnaire was filled in by each participant.

Results

The assessment of GO-DSS was conducted according to the evaluation protocol by 12 testers with different professional backgrounds, and from different organizations. Five were general practitioners (GPs), two were medical specialists, all of whom were actual experimented practitioners (more than 20 years of practice). One was a pharmacist, two were non-clinician e-health informaticians, and two were engineers. Among them, 66% (n=8) already used a CDSS, and 75%

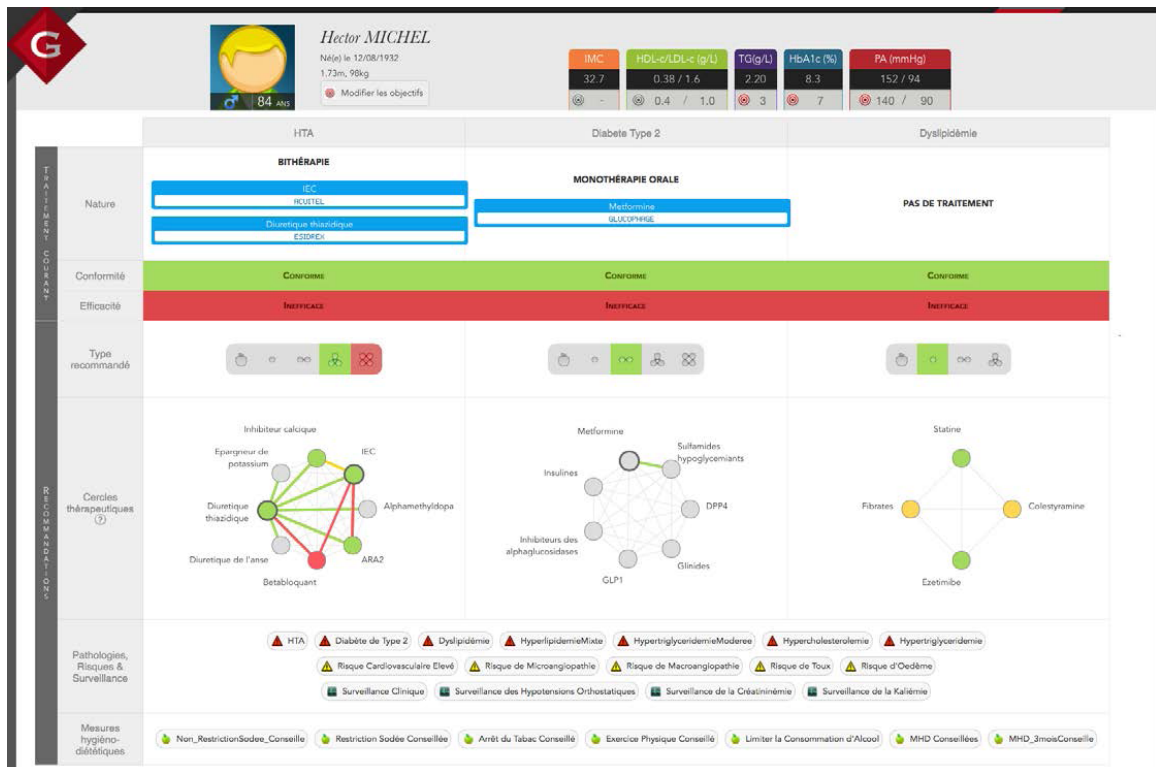


Figure 2 – GO-DSS display of guideline-based recommendations for hypertension, type 2 diabetes, and dyslipidemia

(n=9) were familiar with CPGs. Figure 3 reports the distribution of the participants' answers to the qualitative questionnaire. Most of the responses, beyond 9/12, are positive ("yes" or "rather yes"). Free comments have been manually classified as positive or negative. Positive remarks were: clarity of the user interface, intuitive use, synthetic presentation of the patient case, possibility to get back to the original text of CPGs, the innovative presentation of recommendations as therapeutic circles (sometimes with some reservations). Negative remarks were: lack of connection with any EHR or CPOE, risk of an overloaded interface in case of many comorbidities, need for more classical textual expression of recommendations. The SUS score obtained from the 12 participants was measured as 90.7%. This result, according to the ratings of the standard with respect to usability, classifies the GO-DSS as an "excellent system" of "grade A".

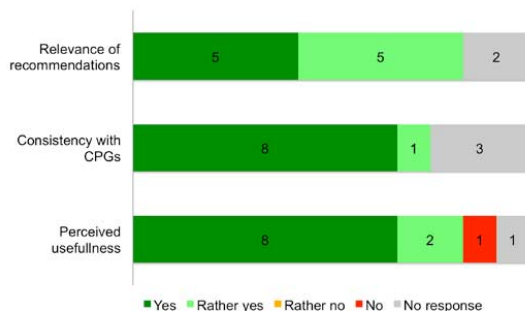


Figure 3 – Distribution of responses to the qualitative questionnaire

Discussion and Conclusion

GO-DSS is a CDSS to manage patients with multiple pathologies from single-disease guidelines. An ontological reasoning process allows for the management of intra and inter CPGs conflicts. Conflict resolution is completed before the display of therapeutic propositions, categorized as possible, recommended, or contra-indicated options. Therapeutic propositions are displayed per pathology in a new graphical representation called "therapeutic circles". GO-DSS processing and interface have been evaluated by a sample of testers on simulated cases. The therapeutic options proposed by GO-DSS have been considered as medically sound by the clinicians, and in good coherence with CPG contents. The innovative presentation of recommendations as therapeutic circles was diversely appreciated: some testers were not comfortable with such a display considering it was not easy to understand, whereas others were enthusiastic asking for some refinements such as the possibility to link GO-DSS to a drug data base which allows to access the list of drugs when clicking on the disc of a given drug class, and select the appropriate drug. This could incrementally build the prescription in a way similar to the "Add to Cart" button used to build a shopping list. All testers considered that with some training, the display of CPGs as therapeutic circles was astoundingly modern and convenient. Another issue discussed concerned the increase in the number of pathologies and the possibility to have comorbidity-based columns used as tabs that could be dynamically opened/closed. The question of having a unique therapeutic patient-centered circle was set but considered as rapidly unreadable.

The design of GO-DSS's interface and the display of medication recommendations as therapeutic circles were finally well accepted by the testers. However, due to the small size of the sample (although the analysis was qualitative), and the fact

that testers worked on simulated cases outside the actual clinical workflow, we cannot conclude that our results may be generalizable, and further work is needed to assess under real conditions the true value of such a display of guideline-based therapeutic recommendations. The first step will be to increase the size of the sample and organize an online evaluation of GO-DSS based on the same evaluation protocol and using GP social/professional networks for recruiting participants.

References

- [1] Alabousi M, Abdullah P, Alter DA, Booth GL, Hogg W, Ko DT, Manuel DG, Farkouh ME, Tu JV, Udell JA. Cardiovascular Risk Factor Management Performance in Canada and the United States: A Systematic Review. *Can J Cardiol*. 2016 Jul 18. pii: S0828-282X(16)30208-2.
- [2] Wennberg JE. Forty years of unwarranted variation—and still counting. *Health Policy* 2014;114(1):1–2.
- [3] Shojania K, Jennings A, Mayhew A, Ramsay C, Eccles M, and Grimshaw J. Effect of point-of-care computer reminders on physician behaviour: a systematic review. *CMAJ* 2010;182(5): 216–25.
- [4] Jaspers MW, Smeulders M, Vermeulen H, Peute LW, Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings, *J Am Med Inform Assoc* 2011;327–334.
- [5] Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. *Implement Sci* 2012;7:50.
- [6] Shiffman RN, Liaw Y, Brandt CA, and Corb GJ. Computer-based guideline implementation systems: a systematic review of functionality and effectiveness. *JAMIA* 1999; 6 (2): 104–14.
- [7] Roshanov PS, Fernandes N, Wilczynski JM, Hemens BJ, You JJ, Handler SM, et al. Features of effective computerised clinical decision support systems: meta-regression of 162 randomised trials. *BMJ* 2013;346:f657
- [8] Séroussi B, Bouaud J. Reminder-based or on-demand decision support systems: a preliminary study in primary care with the management of hypertension. *Stud Health Technol Inform*. 2004;101:142–6.
- [9] Hughes LD, McMurdo ME, Guthrie B. Guidelines for people not for diseases: the challenges of applying UK clinical guidelines to people with multimorbidity. *Age Ageing*. 2013 Jan;42(1):62–9.
- [10] Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* 2012;380:37–43.
- [11] Blozik E, van den Bussche H, Gurtner F, Schäfer I, Scherer M. Epidemiological strategies for adapting clinical practice guidelines to the needs of multimorbid patients. *BMC Health Serv Res*. 2013 Sep 16;13:352.
- [12] Wilk S, Michalowski W, Michalowski M, Farion K, Hing MM, Mohapatra S. Mitigation of adverse interactions in pairs of clinical practice guidelines using constraint logic programming. *J Biomed Inform*. 2013 Apr;46(2):341–53.
- [13] Michalowski M, Wilk S, Rosu D, Kezadri M, Michalowski W, Carrier M. Expanding a First-Order Logic Mitigation Framework to Handle Multimorbid Patient Preferences. *AMIA Annu Symp Proc*. 2015 Nov 5;2015:895–904.
- [14] Abidi SR, Abidi SS, Hussain S, Shepherd M. Ontology-based modeling of clinical practice guidelines: a clinical decision support system for breast cancer follow-up interventions at primary care settings. *Stud Health Technol Inform*. 2007;129(Pt 2):845–9.
- [15] Daniyal A, Abidi SR, Abidi SS. Computerizing clinical pathways: ontology-based modeling and execution. *Stud Health Technol Inform*. 2009;150:643–7.
- [16] Shortliffe, E. H. and Cimino, J. J. (Editors), *Biomedical Informatics: Computer Applications in Healthcare and Biomedicine*: 4th Edition, Springer, New York (2013).
- [17] Cardoso RL, Leite RO, de Aquino AC. A Graph is Worth a Thousand Words: How Overconfidence and Graphical Disclosure of Numerical Information Influence Financial Analysts Accuracy on Decision Making. *PLoS One*. 2016 Aug 10;11(8):e0160443.
- [18] Shaban-Nejad A, Lavigne M, Okhmatovskaia A, Buckridge DL. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Ann N Y Acad Sci*. 2016 Oct 17.
- [19] Wipfli R, Ehrler F, Bediang G, Bétrancourt M, Lovis C. How Regrouping Alerts in Computerized Physician Order Entry Layout Influences Physicians' Prescription Behavior: Results of a Crossover Randomized Trial. *JMIR Hum Factors*. 2016 Jun 2;3(1):e15.
- [20] Payne TH, Hines LE, Chan RC, Hartman S, Kapusnik-Uner J et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. *J Am Med Inform Assoc*. 2015 Nov;22(6):1243–50.
- [21] Griffon N, Kerdelhué G, Hamek S, Hassler S, Boog C, Lamy JB, Duclos C, Venot A, Darmoni SJ. Design and usability study of an iconic user interface to ease information retrieval of medical guidelines. *J Am Med Inform Assoc*. 2014 Oct;21(e2):e270–7.
- [22] Galopin A, Bouaud J, Pereira S, Séroussi B. An Ontology-Based Clinical Decision Support System for the Management of Patients with Multiple Chronic Disorders. *Stud Health Technol Inform*. 2015;216:275–9.
- [23] Galopin A, Bouaud J, Pereira S, Séroussi B. Clinical practice guidelines consistency for patients with multimorbidity: a case-study in the management of type 2 diabetes and hypertension. *Stud Health Technol Inform*. 2015;210:344–8.
- [24] Lloyd BW, Barnett P. Use of problem lists in letters between hospital doctors and general practitioners. *BMJ*. 1993 Jan 23;306(6872):247.
- [25] Brooke J. SUS: a "quick and dirty" usability scale. In Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, *Usability Evaluation in Industry*. 1996. London: Taylor and Francis.

Address for correspondence

Brigitte Séroussi
Département de Santé Publique – Hôpital Tenon
4, rue de la Chine
75020 Paris, France
brigitte.seroussi@aphp.fr

Development of a Service-Oriented Sharable Clinical Decision Support System Based on Ontology for Chronic Disease

Yong Shang, Yu Wang, Ling Gou, Chengkai Wu, Tianshu Zhou, Jing-song Li

Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou Zhejiang Province, China

Abstract

Clinical decision support systems (CDSSs) have been proved as an efficient way to improve health care quality. However, the inflexibility in integrating multiple clinical practice guidelines (multi-CPGs), the mass input workload of patient data, and the difficulty in system sharing become barriers of CDSSs implementation. In this paper, we proposed a framework of CDSS for chronic disease based on ontology and service-oriented architecture (SOA) to improve these defects. We used ontology for knowledge base construction on multi-CPGs integration to overcome their differences as well as reduce the input procedure of patient data by ontology reasoning. Furthermore, we built the CDSS on an SOA structure to provide flexibility in system and data sharing, such that patients could get suggestions from the same system for self-management of chronic disease. A typical case was used to validate the CDSS functions and accuracy. Two clients were developed to illustrate the SOA superiority.

Keywords:

Clinical Decision Support Systems; Biological Ontologies

Introduction

A study of doctors' workload in China indicates that on average clinicians work 54 hours a week and provide medical care for 22 outpatients every day [1]. Heavy workload affects health care efficiency and quality. Clinical decision support systems (CDSSs) could offer convenience for clinicians' daily work, provide suggestions based on patient status, and further identify medical errors in clinical practice [2]. There have been studies suggesting that CDSSs can be an efficient instrument for the improvement of medical care [3]. The implementation of CDSSs could lower the workload of clinicians, leading to better health care quality and lower cost of medication. It could also be used for patients' self-management by providing suggestions. Despite the advantages of CDSSs, their implementation still faces challenges. Some important considerations are: (1) proper methods for the representation of knowledge, (2) reductions on patient data input, and (3) function integration, reuse, and sharing between multiple medical systems [3][4].

CDSSs often derives knowledge from Clinical practice guidelines (CPGs). CPGs are systematically developed instruments to provide references at specific clinical situation [5]. However, patients may get through multiple disorders, which requires the combination of heterogeneous CPGs. The concepts between CPGs may have both common and specific disease medical knowledge. Also, since CPGs are mainly text-based, they may describe a concept in different syntax and

even in different languages. To illustrate decision support properly according to multi-CPGs, the knowledge base must have a formalized structure to integrate concepts between CPGs and a standardized representation of concepts from different CPGs.

Ontology can be used as a suitable approach for multi-CPGs integration. Ontology can describe and structure knowledge in a formalized and standardized way. Heterogeneous CPGs can be extracted into one ontology model, and be represented as concepts and attributes in a standardized form. Knowledge between CPGs can be integrated properly using this ontology method.

Furthermore, many CDSSs are developed with specific computing requirements. Cross-platform adoption of CDSSs faces obstacles; data and suggestions from CDSSs are hard to share by different users such as in clinician-patient communication. This means a great waste to the knowledge base and functions in a CDSS. Also, with the development of medical knowledge, continuous improvement of knowledge and methods is a key to CDSSs [6]. Thus the quality of decision supports increases synchronously if the upgrade of a CDSS can instantly consider various users.

The service-oriented architecture has the potential to overcome the difficulty. A SOA is designed to allow different platforms to use services and create a new work process [7]. Applying SOA to a CDSS can modularize the system into various services that can be easily shared by different users. Developers can integrate required services and create personalized CDSS clients for different users, reducing the difficulty of integration through the interfaces of CDSS services. Clinicians and patients using the same CDSS service could get suggestions simultaneously in order to improve patients' self-management and clinician-patient communication.

The main purpose of this study is to design a service-oriented CDSS base on ontology. The system shall provide the flexibility in the integration of multi-CPGs and the sharing of CDSS system.

Methods

Type 2 Diabetes Mellitus and Hypertension CPGs

We apply CPGs of Type 2 diabetes mellitus (T2DM) and hypertension (HT). The patient group of these disease is vast and these diseases are contribute to a large medication cost around the world. Also, T2DM and HT require patients' self-management to improve prognosis. We collected T2DM CPGs from the Chinese Diabetes Society (CDS) and the American Association of Clinical Endocrinologists (AACE), and HT

CPGs from the Ministry of Health of the PRC and the Eight Joint National Committee (JNC8).

Knowledge Base Construction

The basic design of the knowledge base used in the system is consist of three parts: (1) CPG knowledge base class, (2) coding system class, and (3) patient class. The CPG knowledge base class is the core function, which stores all the knowledge extract from various CPGs. The coding system class is used to store codes that are used in hospital and medical institutes, like ICD-9-CM codes, for data mapping and auto input. The patient class is to create patient individuals with clinical data. The whole knowledge base was built by Protégé and encoded in OWL2.

The formalization representations of disease knowledge are manually extract and standardized from T2DM and HT CPGs. As mentioned, CPGs between T2DM and HT may contain intersection concepts and disease special concepts. Thus the main goal of the knowledge base modeling is to create infrastructure ontology and special ontology. The infrastructure ontology contains basic medical concepts for T2DM and HT CPGs, such as factors that cause the disease, treatment that may cure the disease, and examinations that should be done for the disease. Figure 1 shows the infrastructure of the knowledge base. The infrastructure also contains properties to give relations between individuals in these classes. For example, the property *hasContraindication* link individuals in *Medicine class* and *Disease class*, to express the relation when a drug is contraindicated to a certain disease. These properties will be used when filling the classes with detail individuals as special ontology. Figure 1 only shows part of the properties.

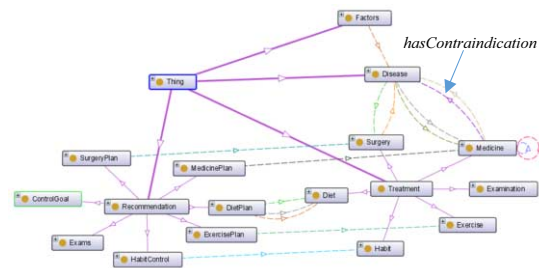


Figure 1 – Knowledge base infrastructure

The special ontology contains more granular disease concepts that are mainly used by specific disease. These concepts will be treated as individuals to fill the infrastructure classes. For examples, in *Factors class of the special ontology*, we extracted disease causing factors from CPGs as individuals to fill under the class, and created a relationship using *hasFactor* property to link the factors to certain disease. By using OWL (explain this) *Import* property, the construction of special ontology for T2DM and HT can carry on simultaneously. In a new Protégé project, importing the infrastructure ontology and the T2DM disease knowledge base will have the same structure as the HT knowledge base. Since the structure and the properties are the same, it enables cross diseases decision support when the system needs to combine more diseases.

Concepts from different CPGs may be applicable to such ontology knowledge base. The classes of ontology knowledge base are listed in table 1. The treatment class consists of all basic treatment for diseases extract from CPGs. The recommendation, however, is the formalized treatment plan consisting of items in Treatment class to provide complete clinical suggestions from CPGs.

In addition to rule-based concepts extracted from CPGs, the rule system consists of rules for diagnosis, exams, medicines, surgeries, habits, data mapping, and evaluation. The rule system is encoded in Jena Rules and saved as separate files according to different functions. The rule system provides criteria to support a CDSS. Figure 2 shows an example.

```
[Dis_Diabetes1:(?x CDSS:P_FBG ?y)
(CDSS:FPG CDSS:FBG_CriticalLimit ?z)
greaterThan(?y, ?z)
->( ?x CDSS:hasPatientDisease CDSS:Diabetes)]

[DrugPlanUnT1:(?x CDSS:hasPatientDrugRec ?y)
(?x CDSS:hasAllergic ?z)
(?y CDSS:hasPlanDrug ?pd)
equal(?pd,?z)
->( ?x CDSS:hasPatientDrugUnRec ?y)]
```

Figure 2 – Example rule

The coding system class is used for integrating various systems. It is helpful when mapping patient data into a CDSS. In hospitals, the diagnosis and exam results have codes to indicates their categories. The code are automatically filled in the patient data to reduce manual work when using the ontology system. When inputting patient data with codes, the reasoner compares the code in the data to the code in the ontology, and maps the same data into patient individuals' attributes.

The patient class is initially empty for users to create their own patients for decision support, which consist of exam results, diagnosis, daily behaviors, etc. These data are fundamental for CDSS to give personalized recommendation.

Table 1 – CPG Knowledge base structure

Class	Subclass	Description
Disease		Store diseases about T2DM and HT and other complications or contradiction
Factors	DailyBehavior ExamResult BodyState Heredity	Store disease factors that may influence disease diagnosis and suggestion output
Treatment	Diet Exam Exercise Habit Medicine Surgery	Basic treatment measures that may used for disease treatment and management. The items are fully isolated.
Recommendation	ControlGoal DietPlan ExercisePlan HabitControl MedicinePlan SurgeryPlan	Recommendations extract from CPGs and from clinic experts. The items are consist of individuals from Treatment class
Unit		Unit may used in the system

System architecture

To provide flexibility for the integration and reuse of the CDSS, the system was built in the form of a web service. As figure3 illustrated, the system is separate into three parts: (1) knowledge base, (2) CDSS web services, and (3) clients. The CDSS services acquire input such as ontologies and rules from

the knowledge base. They also get patient information and instructions from client programs. This would allow the system to provide personalized recommendation to the clients.

The the knowledge base of the system was implemented using Protégé and Jena Rules . It contains all the medical concepts, properties and rules needed in the system. The knowledge base is maintained by medical experts. Client users communicate with the expert to give feedback so the experts can fine tune the knowledge base. The client users however, can store their frequently used health care plans (e.g. prescription) into the knowledge base for easier reuse. Other types of users may also benefit from the plans stored by the client users, in order to reduce the workload of clinicians.

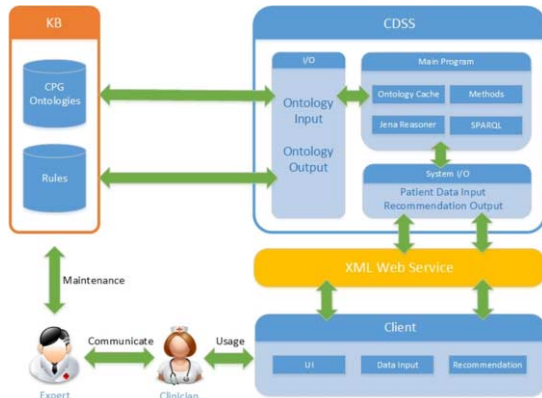


Figure 3 – System Structure

The the main CDSS services of the sytem was built in Eclipse IDE using Java and Jena API, which is a library for ontology developemnt and rule reasoning. This compoennt was further rebuilt in Visual Studio as an ASP.NET Web API to provide RESTful CDSS web services. The CDSS reads in the ontology and rules from the knowledge base. Each user will have their own knowledge base ontology in the server. Since the ontology contains patient class for customized data input, this will keep the patient data secure and private. The cached ontology will be released once a client finished the use, so the patient data will never store in the CDSS system. The functions in the CDSS are encoded in different methods, and clients may select methods to fullfil their own purposes. Under the help of the coding system class, client can input patient data with their category code once for all, and the CDSS will automatically map these data into patient individuals. This could significantly reduce manual work for data input.

The client part was built for demonstration in this study. The structure of the system provides flexibility for client implementation and system integration. Either Ajax or ASP.NET Web API Client or other libraries may be used as a tool for services invoking and all the input and output data model are given as documents. The developers could integrate necessary services into their own program workflow to create a personalized CDSS. Data and suggestions from these CDSSs could share among clients.

Service methods

For the flexibility of invoking and integration, all functions of the CDSS are built as separate methods. Methods are relatively independent so the client may invoke only needed methods and create own CDSS workflow. Each method is encoded in Eclipse IDE in Java and then transfer into ASP.NET Web API with modeled input and output data models. The recommendation functions are mainly achieved

by Jena Reasoner and SPARQL, collecting necessary patient data and medical knowledge to perform ontology reasoning and provide the recommendations.

To provide complete decision support for chronic diseases, the methods are classified into seven categories: 1) Data Input, 2) Knowledge Query, 3) Diagnosis, 4) Exam Suggest, 5) Treatment Suggest, 6) Lifestyle guidance, and 7) Evaluation. Data input methods focus on creating patient individuals for reasoning and providing functions for data mapping. Knowledge query methods mainly provide disease knowledge for referring. Diagnosis, Exam suggest, treatment suggest and lifestyle guidance methods are main functions to support clinical decision through inputted patient data. The Evaluation methods estimate the patients' self-management on chronic diseases and the effectiveness of medicine plans.

System implementation

The CDSS services were hosted on Internet Information Server as a Web API. The sample demonstration clients were developed to invoke methods of CDSS services for decision support. To illustrate the system's functions, we used a typical patient case with T2DM and HT combining with renal insufficiency to acquire clinical recommendation from the CDSS system.

Results

The knowledge base ontology consists of 47 classes and 182 individuals, together with 121 properties and 72 rules. The completeness as well as the correctness of concepts and their relationships were confirmed by clinical experts.

As for the CDSS system, a total of 32 methods were built in Eclipse IDE with Jena API for ontology query and reasoning. These methods were then packed and transferred into Visual Studio for ASP.NET Web API development. The methods were built as a RESTful Web API with formalized data input and output model. The data models were described using document for client development.

Figure 4 – Winform demonstration client diagnosis scenario

A WinForms program and an iOS app were built as demonstration clients to illustrate the invoking and integration of the CDSS services. The WinForms program client performs a complete workflow of decision support progress. It functions as a tool for clinicians to support outpatient service. Figure 4 shows one of the scenarios of the program for disease

diagnosis. Patient data may be inputted into the program, and the program call diagnosis methods from CDSS and give diagnosis suggestions and reasons. Base on that, the user may ask for exam and drug recommendation, diet plan suggestions, and cure effect evaluation from the CDSS as a reference.

The iOS app was much simpler and aimed at diseases self-management for patients. Figure 5 shows an example of the demonstration client. According to the clinician's decision from the WinForms program, the client provides relevant diet plan, exercise plan, daily habit instructions and control goal evaluation. The suggestions and data are shared in the CDSS service. Users may check recommended nutrition, smoke quitting suggestion, as well as weekly exercise time and goals. The evaluation compares patient's current clinical index to the controlled goal and use color coding to alert users on the goals they haven't reached yet.



Figure 5 – iOS app demonstration client scenarios

To illustrate the performance of multi-CPGs integration, we used a typical case of a 44 years old male patient who had T2DM and HT combined with renal insufficiency. The patient data showed FPG at 13.3 mmol/L, total cholesterol (TC) at 7.05 mmol/L, triacylglyceride (TG) at 3.36 mmol/L, HbA1c at 8.5% and a 160/90 blood pressure. The data was inputted into the CDSS by Web API, which returned the diagnosis, exam suggests, medicine suggests and lifestyle suggests. Table 2 shows the suggestions from the system compared to the decision made by a clinician.

The system provided suggestions according to the patient status. These suggestions included diagnosis of HT, T2DM and hyperlipemia based on patient data, and planned routine medical exams for T2DM and HT patient. The clinician made almost the same decision as the system. The system also provided medicine suggestions for T2DM, HT and hyperlipemia. Combining T2DM, HT, renal insufficiency and high TC and TG, the system provided medicines that are safe for patients with renal insufficiency, while maintaining the best choices for T2DM and HT treatment combining medicines for such patient with hyperlipemia like fluvastatin. The medicine chose by the clinician were all included in the suggestions.

Furthermore, contraindicated and cautious medicines were also listed by the system to warn the users, in order to reduce errors on decision making. The system also provided instructions on diet, exercise and habit, while the clinician only provided manuals without detail instructions. It is worth noting that the system suggested 80 mins of exercise per week rather than 150 mins considering the renal insufficiency of the patient.

All of the suggestions are acquired by the WinForms client to help a clinician's work. Instructions on diet, exercise, and habit are also delivered to users through iOS app client for patient self-management since it used the same CDSS services.

Table 2 – Suggestions from the CDSS compared to the clinician decisions

Object	Suggestions from CDSS	Clinician decision
Diagnosis	T2DM, HT	T2DM, HT
Exams	Hyperlipemia OGTT, HbA1c Blood Pressure Blood routine Test Urine routine Test Blood lipid Test Liver function Renal function	Hyperlipemia OGTT, HbA1c Blood Pressure Blood routine Test Urine routine Test Blood lipid Test Renal function
Medicine for T2DM	Insulin Repaglinide Nateglinide Gliquidone Acarbose	Insulin Repaglinide
Medicine for HT	Losartan Valsartan Captopril Nifedipine Metoprolol	Nifedipine Losartan
Medicine for hyperlipemia	Fluvastatin Simvastatin	Fluvastatin
Medicine cautious and contraindicated	Metformin Rosiglitazone Thiazide Aspirin sitagliptin	Not mentioned
Diet	Vitamin B, D Salt < 6g/day Alcohol < 25g/day Cholesterol < 300mg/day Fiber > 14mg/kcal	Provide Education Manual of T2DM and HT
Exercise	80min/week moderate intensity exercise	Tone up more exercise than sitting
Habit	HBGM, HBPM Smoke cessation	Provide Education Manual of T2DM and HT

Discussion

In this study, we adopt ontology methods to develop a multi-CPGs knowledge base to achieve a service-oriented architecture for CDSS services establishment. We built a fully functional CDSS web services as ASP.NET Web API, and developed two demonstration clients to illustrate the invoking of the CDSS services. We also demonstrated the functions of

the CDSS by a typical patient case, which received patient-specific suggestions for treatment decision.

Through the proper integration of CPGs using the ontology methods and the rule system for reasoning, the CDSS provided the best choices for treatment for this patient and gave warning about the contraindicated T2DM and HT medicines due to renal insufficiency of the patient. Since the CDSS functions as web services, the suggestions were able to deliver to both Winform client and iOS app client, leading to timely messages to both clinicians and patients and potential improvement of the communication between clinicians and patients.

Our ontology methods provided great flexibility for the integration between multi-CPGs. We used different editions of T2DM CPGs and HT CPGs to build the CDSS knowledge base. The infrastructure and the special ontology structure made it possible to overcome the difficulties of CPG integration. Those common concepts were captured and formed the infrastructure ontology, to provide OWL classes for both T2DM and HT to use. Disease specific concepts were treated as special ontology to provide knowledge that may be used in certain disease situations. Since ontology aims at formalizing and standardizing description of knowledge [8], the syntax and language differences between CPGs could be overcome. Concepts, relationships, and rules were integrated well into one knowledge base for CDSS to provide suggestions. This may also be extended to other diseases by integrating their CPGs into our knowledge base.

Furthermore, the coding system in the knowledge base reduced the workload when using a CDSS system. Research had indicated that time to use the CDSS system should be minimized [9]. For ontology reasoning, the patient data from data bases can be automatically mapped to the patient individuals, since the data usually contain a code to indicate their category. With this function, the input procedure could be much more efficient. Users could load all the patient data and input into the CDSS once for all instead of classifying the data manually in advance.

By adopting the service-oriented structure, the functions of the CDSS are provided as web services, and can be easily utilized by various platforms. Building a CDSS as web services may reduce workload of client development and system integration. The knowledge and functions in the CDSS can then be shared to various users through service invoking. In this study, we had demonstrated two sample clients, a Winform program and an iOS app, and they functioned properly through different platforms and languages. By utilizing different methods, developers can create personalized clients. Dividing the functions into various sets provides flexibility for developers to integrate decision support into clinicians' workflow, providing decision support at a right time and in a proper way.

It is possible to further improve the CDSS. Since not all the patient data are formalized when stored in data base, patients' free text records still demand manual classification and input. Such manual work may negatively impact the implementation of CDSS and may cause human error. To overcome this issue, natural language processing may be a viable way to formalize free text records and extract useful information through ontology reasoning to map patient data [10][11]. Also, many EMR systems use international standard for data transfer such as HL7. Using HL7 may further lower the barrier in integration and data mapping of our system. Currently the system was only validated by one typical patient case. In the future, we will use larger group of patient data to validate and improve our system.

Conclusions

In this paper, we proposed an ontology and service-oriented architecture to provide flexibility in multi-CPGs integration and CDSS reuse and share, as well as procedure simplification in data input. A CDSS web service was built to provide clinical decision support through ontology reasoning based on a knowledge base derived from T2DM and HT CPGs. Demonstration clients were also developed to show the flexibility in CDSS services invoking, integration, and reuse. These clients achieved various functions by invoking different methods in the CDSS service. In future, we will incorporate natural language process and HL7 standard for data input, and perform cross-ontology reasoning by including more ontologies into the system.

Acknowledgements

This work was supported by Chinese National High-tech R&D Program (No.2015AA020109), National Key Scientific Instrument and Equipment Development Project (No.2016YFF0103200), and the Fundamental Research Funds for the Central Universities of China.

References

- [1] J. Wen, T.Y. Hao, and X.Y. Hu, Doctors' workload in China: A status-quo study, *Chinese Journal of Evidence-Based Medicine* **15** (2015), 133-136.
- [2] D.W. Bates, M. Cohen, L.L. Leape, J.M. Overhage, M.M. Shabot, and T. Sheridan, Reducing the frequency of errors in medicine using information technology, *Journal of the American Medical Informatics Association* **8** (2001), 299-308.
- [3] M.W. Jaspers, M. Smeulers, H. Vermeulen, and L.W. Peute, Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings, *Journal of the American Medical Informatics Association* **18** (2011), 327-334.
- [4] A. Miller, B. Moon, S. Anders, R. Walden, S. Brown, and D. Montella, Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research, *International Journal of Medical Informatics*, **84** (2015), 1009-1018.
- [5] A.M. Audet, S. Greenfield, and M. Field, Medical practice guidelines: current activities and future directions, *Annals of Internal Medicine* **113** (1990), 709-714.
- [6] J.A. Osheroff, J.M. Teich, B. Middleton, E.B. Steen, A. Wright, and D.E. Detmer, A roadmap for national action on clinical decision support, *Journal of the American Medical Informatics Association* **14** (2007), 141-145.
- [7] S.R. Loya, K. Kawamoto, C. Chatwin, and V. Huser, Service oriented architecture for clinical decision support: A systematic review and future directions, *Journal of medical systems* **38** (2014), 1-22.
- [8] T.R. Gruber, A translation approach to portable ontology specifications, *Knowledge acquisition* **5** (1993), 199-220.
- [9] E. Kilsdonk, L.W. Peute, and M.W. Jaspers, Factors influencing implementation success of guideline-based clinical decision support systems: A systematic review and gaps analysis, *International Journal of Medical Informatics* **98** (2017), 56-64.
- [10] D. Demner-Fushman, W.W. Chapman, and C.J. McDonald, What can natural language processing do for clinical decision support?, *Journal of biomedical informatics* **42** (2009), 760-772.
- [11] D. Mendes, I.P. Rodrigues, C.F. Baeta, and C. Solano-Rodriguez, Ontology driven controlled natural language clinical decision support system for the cardiovascular specialty, *Procedia Technology* **16** (2014), 1493-1501.

Address for correspondence

Corresponding author: Jing-song Li.

Address: College of Biomedical Engineering and Instrument Science, Zhejiang University, 38 Zheda road, Hangzhou 310027, China.

Email: ljs@zju.edu.cn.

Tel.: +86-571-87951564

The Paradox of Higher Charges for Lower-Risk Inpatient Admissions: When Healthier Patients Cost More

Lincoln Sheets, MD^{a,b}, Kayson Lyttle, BSc^b,

Lori L. Popejoy, PhD, APRN, GCNS-BC^c, Jerry C. Parker, PhD^b

^a Informatics Institute, University of Missouri, Columbia, Missouri, USA

^b School of Medicine, University of Missouri, Columbia, Missouri, USA

^c Sinclair School of Nursing, University of Missouri, Columbia, Missouri, USA

Abstract

Risk stratification is essential to achieving the Triple Aim of better health, better care, and lower costs. Although risk tiers based on chronic disease diagnoses and recent healthcare utilization were predictive of healthcare utilization and charges in a managed population, their correlation with specific high-cost outcomes was unknown. More detailed analyses were performed to confirm that admissions for higher-risk patients were more expensive. However, these analyses found that charges for admissions of high-risk patients were actually not more expensive but 33% less expensive. The billing categories of implants, surgery, and supplies accounted for 93% of this difference. These findings may reflect that high-risk patients are less often appropriate candidates for elective surgery. An understanding of this difference, especially if validated by claims data and replicated in other populations, may lead to important insights into using risk stratification for predicting health services utilization in managed care populations.

Keywords:

Risk Assessment, Practice Management, Health Services Needs and Demand

Introduction

The Chronic Care Model and Risk Stratification

The Chronic Care Model [1] proposes improving the effectiveness of interactions between patients and providers as a way of promoting the “Triple Aim” of healthcare [2]: (a) better health, (b) better care, and (c) lower costs. Evidence has shown that patients with care coordinators have fewer emergency department and urgent care episodes [3], hospital admissions [4], and readmissions [5].

Unfortunately, healthcare often fails to provide effective coordination of care across a target population [6,7]. By bridging the implementation gaps in the Chronic Care Model, well-designed risk stratification supports the transition from the traditional “reactive” model of medical care [8] to proactively planning care in order to maintain health and avoid preventable conditions. Risk stratification is a potentially powerful tool for managing patient risk and improving outcomes, and stratifying risk by previous healthcare utilization has been shown to be a useful predictor of future healthcare needs [9].

LIGHT² and Risk-Stratification Tiers

LIGHT² (Leveraging Information Technology to Guide Hi-Tech and Hi-Touch Care) was a Health Care Innovation Award from the Centers for Medicare and Medicaid Services to examine the use of advanced health information technology and care coordination in a managed population [10]. With over 10,000 patient cases to be managed by fewer than 25 advanced practice nurses, the LIGHT² project needed clinical decision support to help focus preventive and longitudinal care on those patients who were at highest risk. The LIGHT² investigators defined four risk tiers as (1) Healthy, (2) Stable, (3) Unstable, and (4) Complex, on the basis of diagnoses that are included in the Chronic Conditions Data Warehouse (CCW) [11] and on healthcare utilization in the preceding year (see Figure 1).

These risk tiers supplied highly useful information about the prospective risk of patient healthcare utilization and charges [12]. However, their correlation with population-wide average healthcare charges was insufficient to demonstrate their correlation with specific high-cost outcomes such as inpatient admissions. Therefore, a detailed analysis of financial reporting data for the LIGHT² participants was needed to examine the assumption that admissions for high-risk patients were more expensive than for low-risk patients.

Objective

The primary objective was to test the hypothesis that admissions for higher-risk “unstable” and “complex” patients are more expensive than for lower-risk “healthy” and “stable” patients.

Methods

Data Sources

The LIGHT² program enrolled primary care patients at the University of Missouri Health System who were already enrolled in Medicare or Medicaid. The study cohort was comprised of 9,568 patients who were enrolled in LIGHT² on or before July 1, 2013. All data on patient diagnoses, prescriptions, and other clinical attributes were based on the electronic health records of the University of Missouri Health System as maintained by clinicians between 2011 and 2014. Charges per admission for thirteen billing categories were obtained from the IBM Cognos financial reporting system. The billing system is used by the hospital system for accounting and internal fiscal analyses. No missing data were detected.

Definitions

The four risk tiers were defined as shown in Figure 1. Patients who had none of the 27 chronic conditions identified in the Chronic Conditions Warehouse (CCW) [11] constituted the lowest-risk or “Healthy” category (Tier 1). Patients with one or more chronic conditions were placed in higher tiers depending on the number of their outpatient clinic visits and hospital episodes during the year prior to analysis. Patients with five to twelve related outpatient visits or one related hospital episode in a year were defined as “Unstable” (Tier 3), based on the investigators’ clinical judgement. Patients with chronic conditions but fewer visits and episodes than “Unstable” patients were defined as “Stable” (Tier 2), and those with more visits or more episodes than “Unstable” patients were defined as “Complex” (Tier 4).

In order to analyze the relationships between risk tiers and inpatient utilization charges, the hospital charges for inpatient admissions of patients in each tier were summed for each fiscal year. Only admissions for which the primary diagnosis was one of the 27 chronic conditions in the CCW were included in the sum of charges. All charges were measured in US dollars. For each risk tier in each fiscal year, the charges for all chronic-condition-related hospital admissions were averaged, in total and by billing category. Averages by risk tier and fiscal year were then compared, and the differences were subtotaled by billing category.

Results

The average charges by risk tier for chronic-condition-related inpatient admissions in fiscal year 2013 (FY13) and fiscal year 2014 (FY14) are shown in Table 1 and Figure 2. Additionally, Table 1 and Figure 3 subtotal these charges by thirteen billing categories.

In both fiscal years, average charges for chronic-condition-related inpatient admissions of “Complex” (Tier 4) patients were near the lowest for all risk tiers, at US \$10,046 and \$10,123 for the two years. The only lower average charges were for “Healthy” (Tier 1) patients at \$8,162 and \$7,164. The highest average charges were for “Stable” (Tier 2) patients at \$11,954 and \$14,835, and “Unstable” (Tier 3) patients at \$13,633 and \$12,423 (see Table 1 and Figure 2).

In all risk tiers for both fiscal years, nursing is the largest single billing category. However, the difference in average nursing charges for “Complex” patients (\$4431) and Tiers 2 and 3 patients (\$4605 combined average) is only 4% (see Table 1 and Figure 3). Charges in the implants, surgery, and supplies billing categories accounted for 93% of the difference between “Complex” patient charges and charges for “Stable” and “Unstable” patients (see Table 1 and Figure 3).

Discussion

The most remarkable finding of these analyses is that, in both fiscal years, charges for admissions of “Complex” patients were less expensive on average than those for “Stable” (33% higher overall) and “Unstable” (30% higher overall) patients (see Table 1 and Figure 2). Most of the difference seems to be explained by “Complex” patients getting less surgery and fewer implants than “Stable” and “Unstable” patients. This may be because complex-care patients are often less appropriate candidates for elective surgery such as joint replacement surgery.

Although joint replacements are known to be a significant driver of increasing healthcare costs [13-16], this study is the first to our knowledge on the interaction between patient risk stratification, surgical services, and the costs of hospital admissions at the level of population health management. Using risk stratification to help predict future healthcare utilization and charges is an important emerging technique in practice management. However, the “bottom line” numbers must be understood in some detail in order to discover the knowledge hidden in the data, as in this counterintuitive finding. Because this study was conducted within a US academic medical center, the findings may not be generalizable to every healthcare setting. However, if the cause of the higher charges incurred by relatively healthier patients is related to greater numbers of high-cost elective surgeries, such as those involving implants, this pattern might be expected in any healthcare system that routinely performs such surgical procedures.

Limitations

The use of charges as a proxy for costs limited precision of the financial analyses, particularly because nursing charges are the largest cost category and are calculated as fixed overhead cost. Since data was collected from a single healthcare system, some utilization of providers outside the system may be missing. This could affect both the risk stratification of patients and the number and cost of their admissions after stratification. These data gaps could be resolved by supplementing electronic medical records with insurance claims data for the same population over the same period.

Future Research

Although the billing categories provide some important insights into the lower charges for “Complex” patient admissions, more detailed billing data is needed to confirm that the type of surgeries and supplies reduced in “Complex” admissions are, indeed, related to implants and that these implants are for elective surgeries for which the complex-care patients were less appropriate candidates. More detailed billing data and chart reviews are also needed to answer what kinds of implants represent most of the difference in the charges, and whether the differences in charges are related to length of stay. Data for additional years should be analyzed to strengthen the association, and this analysis should be validated by replication in other populations.

Conclusion

The finding that “Complex” patients have markedly less expensive inpatient admissions, on average, than “Stable” and “Unstable” patients is contrary to the authors’ original hypothesis, and seems paradoxical. However, it may be explained by “Complex” or high-risk patients receiving less surgery and fewer implants than better surgical candidates in the “Stable” and “Unstable” risk tiers. An understanding of this difference, especially if validated by claims data and replicated in other populations, may lead to important insights into using risk stratification for predicting health services utilization in managed care populations.

Table 1 – Average Charges per Admission (Chronic Condition Related) by Category, US\$

Tier (Year)	Trauma	Surgery	Supplies	Pharmacy	Other	Nursing	Lab	Implants	Imaging	Emergency	Clinic	Cardiology	Ambulance	Total
Tier 1: Healthy (FY13)	0	1,004	626	544	460	3,808	374	603	236	173	8	324	0	8162
(FY14)	0	833	623	502	150	2,671	428	330	118	266	4	1,241	0	7164
Tier 2: Stable (FY13)	11	1,476	914	905	583	3,863	747	2,382	407	307	15	343	0	11,954
(FY14)	0	1,769	1,075	1,185	869	5,144	839	2,747	420	270	12	505	0	14,836
Tier 3: Unstable (FY13)	0	1,533	1,063	1,110	818	4,919	755	2,468	387	267	14	299	0	13,633
(FY14)	3	1,539	944	862	739	4,503	712	2,183	323	282	18	313	0	12,423
Tier 4: Complex (FY13)	0	764	575	1,076	754	4,394	854	615	379	323	27	285	0	10,046
(FY14)	0	856	472	881	665	4,468	909	937	336	357	21	222	0	10,123

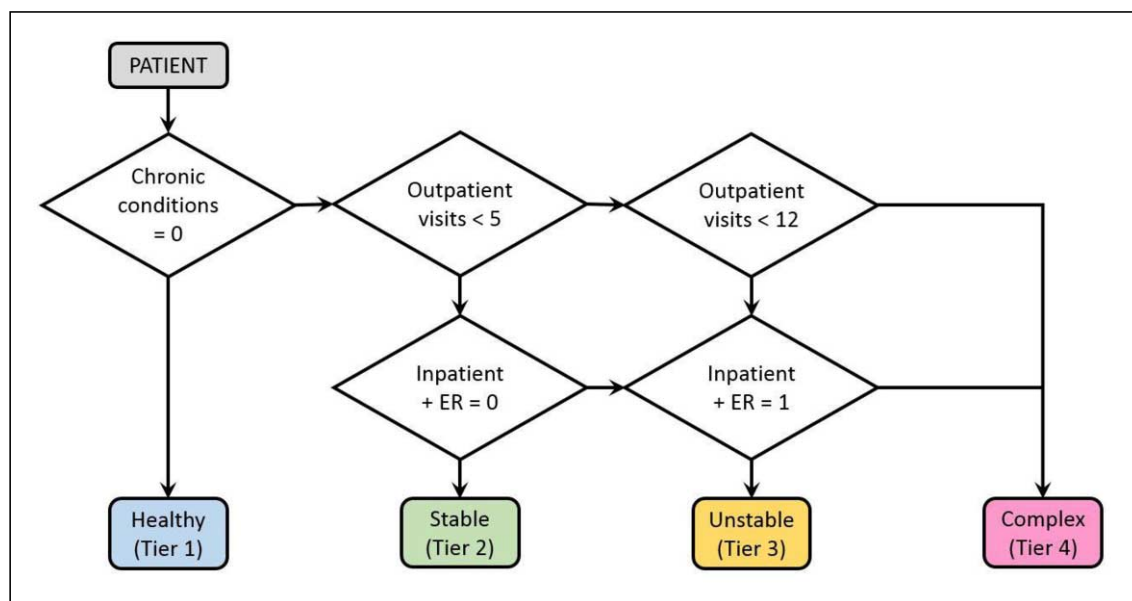


Figure 1 – Definitions of Risk Tiers

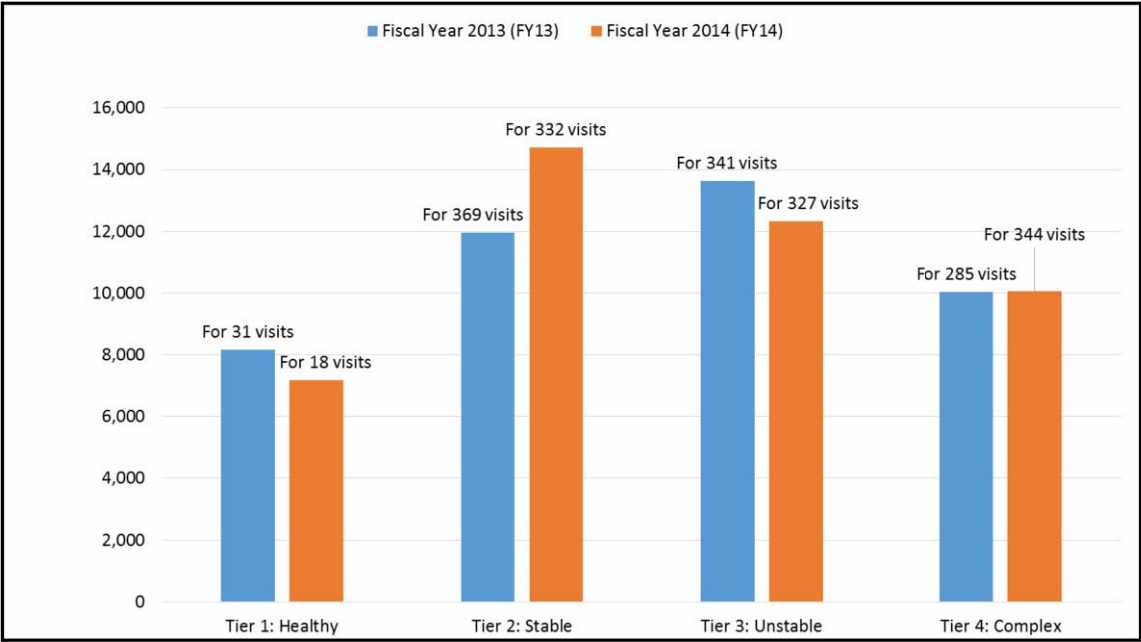


Figure 2 – Average Charges per Admission (Chronic Condition Related), US\$

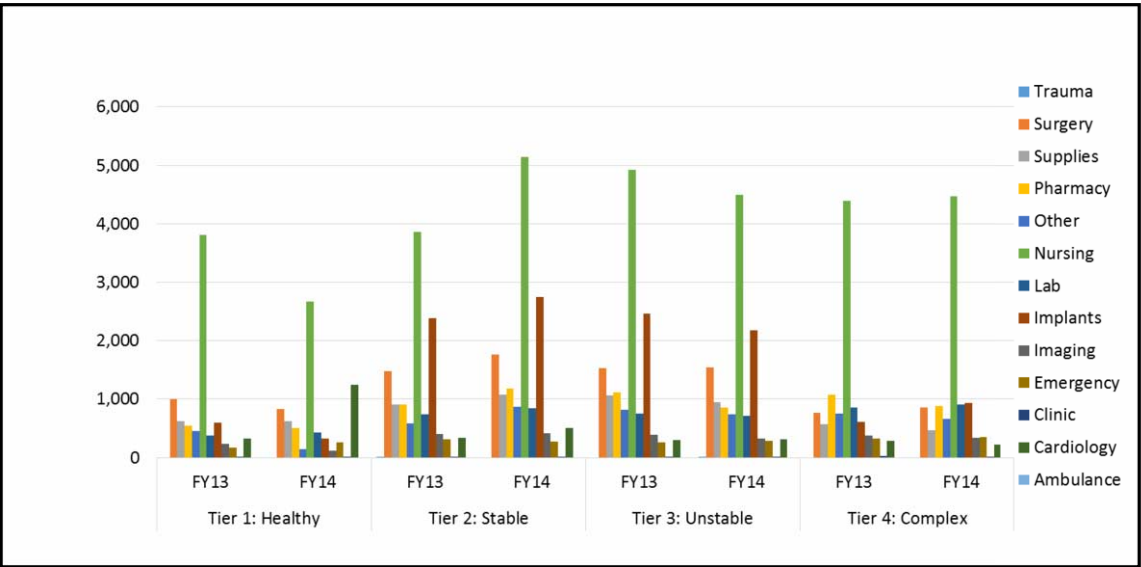


Figure 3 – Average Charges per Admission (Chronic Condition Related) by Category, US\$

Acknowledgements

This publication was made possible by Grant Number 1C1CMS331001-01-00 from the Department of Health and Human Services, Centers for Medicare & Medicaid Services. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. Department of Health and Human Services or any of its agencies. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The research presented here

was conducted by the awardee. Findings might or might not be consistent with or confirmed by the independent evaluation contractor.

The first author, LS, would like to thank his academic advisor, Dr. Chi-Ren Shyu of the University of Missouri Informatics Institute, for advice and guidance in the preparation of this manuscript.

References

- [1] E. H. Wagner, R. E. Glasgow, C. Davis, et al, Quality Improvement in Chronic Illness Care: A Collaborative Approach, *The Joint Commission Journal on Quality Improvement* **27** (2001), 63-80.
- [2] D. M. Berwick, T. W. Nolan, and J. Whittington, The Triple Aim: Care, Health, and Cost, *Health Affairs (Millwood)* **27** (2008), 759-769.
- [3] R. L. Kruse, D. Parker Oliver, D. R. Mehr, G. F. Petroski, D. L. Swenson, and S. C. Zweig, Using Mortality Risk Scores for Long-Term Prognosis of Nursing Home Residents: Caution Is Recommended, *The journals of gerontology Series A, Biological sciences and medical sciences* **65** (2010), 1235-1241.
- [4] C. Sanderson and J. Dixon, Conditions for which Onset or Hospital Admission is Potentially Preventable by Timely and Effective Ambulatory Care, *Journal of health services research and policy* **5** (2000), 222-230.
- [5] K. J. Verhaegh, J. L. MacNeil-Vroomen, S. Eslami, S. E. Geerlings, S. E. de Rooij, and B. M. Buurman, Transitional Care Interventions Prevent Hospital Readmissions for Adults with Chronic Illnesses, *Health Affairs (Millwood)* **33** (2014), 1531-1539.
- [6] T. Bodenheimer, E. H. Wagner, and K. Grumbach, Improving Primary Care for Patients with Chronic Illness: The Chronic Care Model, Part 2, *Journal of the American Medical Association* **15** (2002), 1909-1914.
- [7] K. Coleman, B. T. Austin, C. Brach, and E. H. Wagner, Evidence on the Chronic Care Model in the New Millennium, *Health Affairs (Millwood)* **28** (2009), 75-85.
- [8] R. Snyderman and R. S. Williams, Prospective Medicine: The Next Health Care Transformation, *Academic Medicine* **78** (2003), 1079-1084.
- [9] L. R. Wherry, M. E. Burns, and L. J. Leininger, Using Self-Reported Health Measures to Predict High-Need Cases Among Medicaid-Eligible Adults, *Health Services Research* **49** (2014), 2147-2172.
- [10] L. L. Popejoy, J. Jaddoo, J. Sherman, C. Howk, R. Nguyen, and J. C. Parker, Monitoring Resource Utilization in a Health Care Coordination Program, *Professional Case Management* **20** (2015), 310-320.
- [11] R. A. Goodman, S. F. Posner, E. S. Huang, A. K. Parekh, and H. K. Koh, Defining and Measuring Chronic Conditions: Imperatives for Research, Policy, Program, and Practice, *Preventing Chronic Disease* **10** (2013) 1-16.
- [12] L. Sheets, L. L. Popejoy, M. Khalilia, G. F. Petroski, and J. C. Parker, Identifying Patients at Risk of High Healthcare Utilization, *AMA Annual Symposium Proceedings*, **2016** (2016), 1129-1138.
- [13] S. Kurtz, K. Ong, E. Lau, F. Mowat, and M. Halpern, Projections of Primary and Revision Hip and Knee Arthroplasty in the United States from 2005 to 2030, *The Journal of Bone & Joint Surgery* **89** (2007), 780-785.
- [14] E. Stranges, C. A. Russo, and B. Friedman, Procedures with the Most Rapidly Increasing Hospital Costs, 2004- 2007, Healthcare Cost and Utilization Project (H-CUP) *Statistical Briefs* **82** (2009), 1-9.
- [15] N. A. Wilson, E. S. Schneller, K. Montgomery, and K. J. Bozic, Hip And Knee Implants: Current Trends And Policy Considerations, *Health Affairs* **27** (2008), 1587-1598.
- [16] D. C. Crawford, C. S. Li, S. Sprague, and M. Bhandari, Clinical and Cost Implications of Inpatient versus Outpatient Orthopedic Surgeries: A Systematic Review of the Published Literature, *Orthopedics Review (Pavia)* **7** (2015), 6177.

Address for correspondence

Lincoln Sheets, University of Missouri, Columbia, Missouri, USA
 telephone: 417-860-1197; fax: 573-884-4808

Email: SheetsLR@health.missouri.edu

A Design of a Surgical Site Verification System

Biyu Shen, Yan He, Haoyang Chen

Department of No.2 Affiliated Hospital of Nantong University, Nantong City, Jiangsu Province, China

Abstract

Patient security is a significant issue in medical research and clinical practice at present. The Surgical Verification System (Patent Number: ZL 201420079273.5) is designed to recognize and check surgical sites of patients so as to ensure operation security and decrease the risk for practitioners. Composition: (1) Operating Room Server, (2) Label Reader, (3) E-Label, (4) Surgical Site Display, (5) Ward Client, (6) Label Reader-Writer, and (7) Acousto-Optic Alarm. If the Surgical identification, the surgical site, and so on are incorrect, a flashing label control will appear when the alarm rings. You can specify a sound to play for the alarm, a picture to draw, and a message to send. It is a user-friendly system.

Keywords:

Safety Management; Medical Errors

Introduction

The operating room is an essential place for the operation, medical exam, diagnosis and rescue in a hospital. As a result, surgical care in the operating room has a direct effect on patients' security and recovery.

Patient security is a significant issue in medical research and clinical practice at present. It is pointed out in the Principal Measure of Patient Safety Standard of CHA2017 issued by the Chinese Hospital Association (CHA) in 2016 that hospitals should effectively ensure the safety of patients, making an accurate ID recognition and taking strict measures to avoid mistakes in the identification of patients, operation sites and operational methods [1]. A project named Safe Surgery Saves Lives was launched by WHO in June 2008, in which a Surgical Safety Checklist was offered to surgical teams [2]. Then in 2009, a Surgical Safety Checklist and Risk Assessment was released by CHA and required that information checking should be fulfilled by surgeons, anesthetists and nurses before the induction of anesthesia, the skin incision, and the patient leaving the operating room [3]. The accurate ID and information related to the identification is the prerequisite of a safe operation. The ID and other related information about patients are ascertained through electrical information technology in operating rooms in developed countries in Europe and America.

Nowadays, the construction of modern operating rooms is oriented towards that of digital ones, which are key components of digital hospital construction. There is usually not enough time to carefully check patient-related information because of the successive and fast-paced operation arrangement. Medical incidents resulting from information mistakes occur from time to time, not only in China, but in other countries [4,5]. Therefore, The Surgical Site Verification System (Patent Number: ZL 201420079273.5) was designed to better ensure the safety of patients and reduce risks of clinicians and health-care professionals.

Methods

Composition: (1) Operating Room Server, (2) Label Reader, (3) E-Label, (4) Surgical Site Display, (5) Ward Client, (6) Label reader-writer, and (7) Acousto-optic Alarm.

When a patient enters an operating room, the Label Reader will scan the E-Label of him/her for ID information and then check whether it is in accordance with that of the operating room information. If it's correct, a body picture with the highlighted surgical site will be seen on the Surgical Site Display. Then the operation will begin when surgeons touch the screen to confirm the site information.

Results

When a patient enters an operating room, the Label Reader will scan the E-Label of him/her for ID information and then check whether it is in accordance with the information stored in the Operating Room Server. If it is correct, a body picture with the highlighted surgical site will be seen on the Surgical Site Display. Then the operation will begin when surgeons touch the screen to confirm the site information. In some applications, the Surgical Site Display is a touch screen, which makes a more convenient and clearer verification by touching the screen. The Surgical Site Verification System also includes an Acousto-optic Alarm, which is linked to the Operating Room Server. It will keep shining and making a sound for information confirmation.

The ID information of the patient is recorded in the E-label and will be recognized by the Label Reader connected with the Operating Room Server. The Surgical Site Display is connected with the Operating Room Server to illustrate the surgical site. The patient is equipped with the E-label and the Surgical Site Display is placed in the operating room. The E-label is a kind of electric chip to store information, which is put in a wrist-band of the patient in case of information losses or mistakes.

Discussion

Guarantee of Accurate Patient Information: Surgical mistakes are one of the most serious incidents particularly in operating rooms. The American College of Surgeons reported that surgical site mistakes are the biggest medical incident in the country and set it as one of the most important goals in the operating room to ensure right patients, right surgical sites and right operating processes [6]. It is the prerequisite of safe medical care to correctly recognize the ID information of patients, which is also a guarantee for medical safety.

The ID wrist band contributes to a safe hospital management in the right direction, and offers an advanced and reliable auxiliary device in standard hospital management. The E-label could store accurate and reliable information of patients and help practitioners quickly identify the patient's information. It contributes to a careful, definite and labeled information check. If

there are any mistakes, the system will give a warning as an instant supervision. In contrast, a traditional wrist band is usually written by hand which has problems of unclear handwriting and possibly wrong information. The employment of the E-label involves double-checking by both human and machine, which increases the security of the operation. Especially for children, the elderly, and coma patients, it can greatly increase their security.

Reinforcement of Team Cohesion: The identification of patients and verification of surgical sites need to be implemented in various aspects and different times. People in clinical, technical and logistic departments are involved including patients, ward nurses, operating nurses, anesthetists and surgeons. Coordination and cooperation are necessary to make a successful treatment. Precaution is the most important point in safety management in the operating room. As a result, it's necessary not only to introduce advanced technology but to lay emphasis on the team building. A surgical team consists of surgeons, anesthetists and operating nurses and have the main power to fulfill the surgery treatment. An excellent team is an efficient, optimized group of members who have complementary skills and skills so as to form an integral resultant force. Operating room workers should strictly obey the regulations, cultivate a rigorous attitude towards work and a high responsibility. Furthermore, effective precautions should be taken into account from the perspective of management and the Surgical Site Verification System is designed for this purpose.

It is estimated that there are 234,000,000 operations every year in the world [7]. According to a survey in 2009, the incidence of postoperative complications decreased from 11.0% to 7.0% with the application of the Surgical Safety Checklist [8], which proved most surgical mistakes could be avoided by safety checking. Nowadays [9], medical care in the world is becoming safer, more quality-oriented and much more economical [10,11], where surgical safety is key. The prerequisite of safety management is the implementation of safety checking regulations and procedures while the precaution of adverse events is the emphasis of operating room safety and risk management.

A mistake will cause results ranging from operation delay, treatment effect, waste of labor to serious incidents including disability and death of patients. The reliable and stable Surgical Verification System, as referred to in this paper, can help reduce the mistakes in operation. The double check by human and machine can avoid potential problems in operation, strengthen cooperation among team members, and increase work efficiency so as to contribute to better economic and social benefits.

The potential limitation is that this is just a design, there is no achievement of transformation. The relevant problems could be solved in a better way through data validation and after the transformation.

We have the 3D technology, and the computer will show the anatomy of the system, if the positioning of the scan is correct, then there will be flashing green light though, if not, then there will be a red light. It will ensure the correct operation site as to ensure a smooth operation.

Conclusion

Identification of Patients: The ward nurses type in the identification information in the Ward Client, which includes ward section, ward bed number, name, gender, age, surgical site and remark, etc. Then the identification of patients will be transferred into the E-label by the Label reader-writer and sent to the operation room server so that the information will be stored for checking later.

Identification of Surgical Sites: The Label Reader is linked with the Ward Client set in the ward area of the Inpatient Department and is also linked with the Operating Room Server to read the information of patients in the E-label. The Surgical Site Display, therefore, could point out the surgical site when connected with the Operating Room Server.

The Ward Client is connected with the Operating Room Server by the China Mobile GSM network. When the patient enters the operating room, the Label Reader will scan the E-Label of him/her for ID information and convey it to the Operating Room Server to check whether it's in accordance with that stored there. If it is correct, a body picture with highlighted surgical site (A) will be seen on the Surgical Site Display and the Acousto-optic Alarm keeps making a warning tone until the surgeon and the other team members observe and touch the screen to confirm the site information (A). Then the operation will begin.

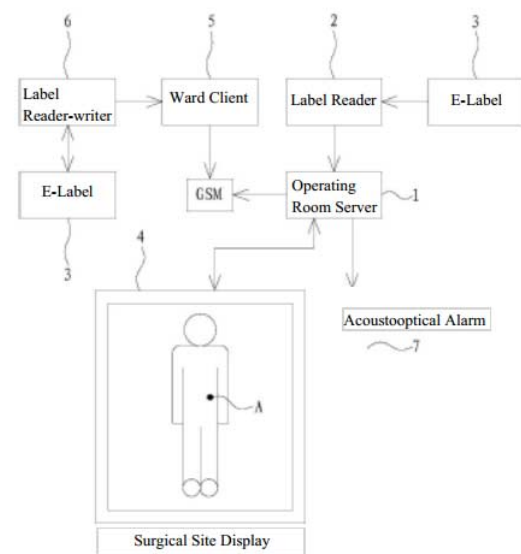


Figure 1 - The Surgical Verification System (Patent Number: ZL 201420079273.5). Composition: (1) Operating Room Server, (2) Label Reader, (3) E-Label, (4) Surgical Site Display, (5) Ward Client, (6) Label Reader-Writer, and (7) Acousto-Optic Alarm.

Acknowledgements

The design has gotten the patent for utility models (ZL201420079273.5). We would like to thank Lingyu Ma, Yan He, Haoyang Chen for their assistance in this study.

References

- [1] Chinese Hospital Association.CHA , Principal Measure of Patient Safety Standard of CHA2017[J]. China Hospital, 2017, 11(1).
- [2] D. F. Milam, Improving patient safety in the operating room: lessons from the aviation industry, Nature Clinical Practice Urology, 2008, 5(6): 289-289.
- [3] X. He, Importance of application of pre-surgical safety check in the operating room[C]. Selection of papers in the 14th Academic meeting of Chinese Nursing Association (Volume one),2010.
- [4] L. Wangoo, R. A. Ray, Y. H. Ho, Attitudes and compliance to the WHO Surgical Safety Checklist; a review. *European Surgery*, 2016, 48(2): 155-162.
- [5] S. J. Singer, G. Molina, Z. Li, W. Jiang, S. Nurudeen, J. G. Kite, L. Edmondson, R. Foster, A. B. Haynes and W. R. Berry, Relationship between

- operating room teamwork, contextual factors, and safety checklist performance. *Journal of the American College of Surgeons*, 2016, **223**(4): 568-580. e2.
- [6] B. M. Gillespie, Moving evidence into clinical practice: Implementing surgical safety checklists in the operating room. *The Queensland Nurse*, 2016, **35**(3): 33.
 - [7] T. G. Weiser, S. E. Regenbogen, K. D. Thompson, A. B. Haynes, S. R. Lipsitz, W. R. Berry and A. A. Gawande, An estimation of the global volume of surgery: a modelling strategy based on available data. *The Lancet*, 2008, **372**(9633): 139-144.
 - [8] E. N. deVries, H. A. Prins, R. M. P. H. Crolla, A. J. den Outer, G. van Andel, S. H. van Helden, W. S. Schlack, M. A. van Putten, D. J. Gouma, M. G. W. Dijkgraaf, S. M. Smorenburg and M. A. Boermeester, Effect of a comprehensive surgical safety system on patient outcomes. *New England Journal of Medicine*, 2010, **363**(20): 1928-1937.
 - [9] I. C. Martin, M. Mason and G. Findlay, A surgical safety checklist. *New England Journal of Medicine*, 2009, **360**(22): 2372-3.
 - [10] E. K. Norton, S. J. Singer, W. Sparks, A. Ozonoff, J. Baxter and S. Rangel, Operating room clinicians' attitudes and perceptions of a pediatric surgical safety checklist at 1 institution. *Journal of patient safety*, 2016, **12**(1): 44-50.
 - [11] E. Litvak, H. V. Fineberg, Smoothing the way to high quality, safety, and economy. *New England Journal of Medicine*, 2013, **369**(17): 1581-1583.

Address for correspondence

BiYu Shen

Tel: 13862937317

Mailing Address: Haier 6th Alley, Chongchuan District,
Nantong City, Jiangsu Province.

E-mail: shenbiyu@126.com

Characterizing Restorative Dental Treatments of Sjögren's Syndrome Patients Using Electronic Dental Records Data

Zasim Siddiqui^a, Yue Wang^a, Payal Makkad^a, Thankam Thyvalikakath^{a,b}

^aDental Informatics Core, Indiana University School of Dentistry, Indianapolis, Indiana, USA

^bCenter for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA

Abstract

Scant knowledge exists on the type of restorative treatments Sjögren's syndrome patients (SSP) receive in spite of their high dental disease burden due to hyposalivation. Increased adoption of electronic dental records (EDR) could help in leveraging information from these records to assess dental treatment outcomes in SSP. In this study, we evaluated the feasibility of using EDR to characterize the dental treatments SSP received and assess the longevity of implants in these patients. We identified 180 SSP in ten years of patients' data at the Indiana University School of Dentistry clinics. A total of 104 (57.77%) patients received restorative or endodontic treatments. Eleven patients received 23 implants with a survival rate of 87% at 40 months follow-up. We conclude that EDR data could be used for characterizing the treatments received by SSP and for assessing treatment outcomes.

Keywords

Dental Records; Sjögren's Syndrome; Dental Implants

Introduction

Sjögren's syndrome (SS) is a chronic autoimmune disorder of exocrine glands, particularly salivary and lacrimal gland, characterized by lymphocytic infiltration of affected gland resulting in the dryness of the mouth and eyes [13]. It is the second most common autoimmune connective tissue disease affecting up to 3.1 million Americans, with approximately 1 in every 70 people affected by primary SS [17; 23; 40]. It is common among middle aged people, with a high prevalence in females (female: male 9:1) [6]. It can occur alone as primary SS or in conjunction with other connective tissue diseases as secondary SS, such as rheumatoid arthritis (RA), systemic lupus erythematosus, and systemic sclerosis [13; 16; 37]. Both primary and secondary SS have similar pathophysiology, signs, and symptoms [7]. The exact etiopathogenesis of SS is unclear and considered to be multifactorial. However, its etiology has been associated with endocrine, genetic and viral factors and alteration in the regulation of cell apoptosis [20; 24; 28]. At present, SS is an incurable disease with symptomatic management options. However recent evidence showed effectiveness of early immunomodulation in limiting disease progression among SS patients [33].

SS patients experience a high caries risk due to reduced salivary flow leading to premature tooth loss despite maintaining good oral health, visiting dentists more frequently, using fluoridated toothpaste, and having more awareness about their disease and oral health [3; 4; 10; 31; 38]. Hyposalivation and early loss of teeth significantly interfere with the individual's normal oral functions such as speaking, chewing, and swallowing thereby compromising their physical, social and emotional quality of

life [5; 15; 26; 29]. These patients often require costly, early life restorative treatments to maintain normal oral functions due to tooth loss [8; 38]. Despite their huge dental disease burden, limited studies exist characterizing SS patients' oral health and dental treatments.

To date, most of the knowledge on SS treatments comes through surveys and interviews on patients' experiences and challenges with maintaining good oral health and receiving dental treatments [2; 10]. Results from these studies report SS patients have difficulties with maintaining good oral health, high caries risk and incompatibility with tooth/tissue supported prosthesis due to mucosal dryness. These challenges with maintaining good oral health highlight the need for clinical research. Very few clinical studies have investigated SS patients' oral health or outcomes from dental treatments received. Lately, implant retained prosthesis are heralded as the treatment of choice to replace lost teeth in patients with SS. However, implant success are inconclusive as they are either case reports or studies with small sample size [9; 19; 30].

The historic use of paper-based records in dental clinics makes retrospective studies of SS difficult. Also, challenges with identifying and confirming SS diagnosis and associated comorbidities were a major barrier in performing clinical studies. The increased adoption of EDR in dental practices offers an opportunity to study the outcomes of various dental treatments among SS patients using EDR data [34].

Studies have shown increased adoption of EDR in both private and academic dental settings in United States and this trend is expected to continue in the future [32; 34; 35]. This trend echoes with the EDR adoption and computer usage in dental practices in other countries including Canada, China, UK, and Brazil [1; 12; 18; 21]. Most countries showing high adoption of EDR used it not only for administrative purposes but also for patient care documentation at the point of care. However, few countries have also shown to be using EDR and computers in dental offices mainly for administrative purposes [1; 12]. In US academic settings, more than 90% of the dental schools document patient care using EDR [32].

Approximately 76% of US independent and group practices use EDR for patient care documentation in 2013 [34] compared to 48% of physician offices having a basic Electronic Health Record in 2014 [14]. Thus, EDRs are a potential data source for clinical research. In this study, we demonstrate the use of EDR data in evaluating dental treatment procedures and outcome for SS patients.

The objective of this study was to determine the feasibility of using EDR to characterize the restorative and endodontic dental treatments for SS and assess the longevity of dental implants placed in patients with SS at the Indiana University School of Dentistry clinics. The long-term objective of our

research is to advance our knowledge of SS and to develop best practice guidelines toward improved oral health and quality of life.

Methods

This study was approved by IRB 1611054551. We retrieved a limited data set of patients seen between January 1, 2005 and October 31, 2016 by performing keyword search for the term "Sjogren" in the EDR. We identified 270 records that contained the term "Sjogren" in the medical history forms, progress/clinical notes, specialty and medical consultation forms, caries risk assessment, and management forms. We used keyword search to identify patients diagnosed with SS as patient's medical and medication histories are typically documented in free-text format or within progress notes.

Two trained dental researchers manually reviewed the clinical notes to identify patients who reported having SS. Unambiguous records stating patients diagnosed with SS were included whereas records only mentioning "Sjogren" as a suspected disease, differential diagnosis, or family history etc. were excluded. Disagreements between researchers were discussed and resolved through consensus.

Next, we retrieved the treatment history of these patients using Current Dental Terminology (CDT) [27; 36] codes that are routinely used to document dental procedures performed in dental practices. We identified the CDT codes related to restorative and endodontic treatment procedures and grouped them into five major treatment types (Table 1). Treatment types were further divided into treatment procedures based on the types of materials (resin-based composite, amalgam), location (maxillary, mandibular), and extent (partial, complete) (Table 2). Each treatment procedure included CDT codes representing that procedure type. For instance, Resin-based composite contains codes: D2330 - D2335 and D2390 - D2394). Amalgam restoration contains codes: D2140, D2150, D2160, and D2161.

We performed descriptive statistics on demographics and treatments. Life tables were constructed to assess survival rates of implant procedures. Two researchers also manually reviewed clinical notes of failed implant records to detect reasons for implant failure.

Results

A total of 180/270 patients were identified with SS. Among them, 165 (91.6%) were female, 11 (6.66%) male, and four patients (2.22%) did not report their gender. These patients had a mean age of 63.75 years (standard error: 1.06 years) with 160 (88.89%) of them being 45 years or older. Among the patients who reported ethnicity, 100 (55.56%) were Caucasian and 13 (7.2%) were African Americans. 61 (33.89%) patients did not report their ethnicity. Only 75 (41.66 %) patients had dental insurance. 120 patients (67%) had a follow-up visit of more than one year. The average follow-up was 5.23 years (SE: 0.32 years).

Tables 1 and 2 show the distribution of treatment types and treatment procedures, respectively. 104 (57.77%) patients received restorative and/or endodontics treatment. These patients received 1,085 different restorative and/or endodontics treatments while the remaining patients received oral examinations with diagnostic procedures, prophylaxis treatment, periodontal therapy, or surgical treatment including tooth extraction. Most common restorative procedures were resin-based composite and amalgam restorations followed by fixed partial denture procedures. 24 patients received 41

complete or partial dentures and 25 patients received 33 endodontic treatments. The mean patient age for patients receiving partial dentures, complete dentures, and endodontic treatments was 66.87, 62.14, and 58.78 years, respectively.

Table 1– Number of treatments received by SS patients

Treatment Types	Number of patients (%)*	Number of Procedures	Mean Patient age (SE)
Restorative	90 (50)	866	61.22 (0.42)
Fixed	41 (22.8)	122	62.61 (0.80)
Partial denture	24 (13.3)	41	66.87 (1.12)
Denture	11(6.1)	23	62.41 (2.19)
Implants	25 (13.9)	33	58.78 (1.88)

*Number of patients >104 due to multiple treatments.

Table 2 - Distribution of treatment procedures in Sjögren's syndrome patients

Treatment Procedures	Number of procedures (%)
Restorative treatment	
Resin-based composite restoration	654 (60.28)
Amalgam restoration	150 (13.82)
Post and core	61 (5.62)
Inlay and onlay	1 (0.09)
Total	866 (79.82)
Fixed partial denture (FPD)	
Total	122 (11.24)
Denture	
Complete denture - maxillary	14 (1.29)
Complete denture - mandibular	4 (0.37)
Partial denture- maxillary	7 (0.65)
Partial denture -mandibular	15 (1.38)
Overdenture	1 (0.09)
Total	41 (3.78)
Implants	
Implants - maxillary	12 (1.11)
Implants - mandibular	11 (1.01)
Total	23 (2.12)
Endodontics treatment	
Root canal treatment (RCT)	27 (2.49)
Retreatment of RCT tooth	4 (0.37)
Apicoectomy	1 (0.09)
Therapeutic pulpotomy	1 (0.09)
Total	33 (3.04)
Total treatment procedures	1,085 (100.0)

Eleven patients received 23 implant treatments, 12 (52.17%) in the maxilla and 11 (47.82%) in the mandible. As shown in Table 3, three implants failed (two in maxilla and one in mandible) making survival rate of approximately 87% during an average follow-up period of 40 months. These 3 implants failed in two patients in the second and fifth months after placement. One implant failed due to osseointegration and the

remaining two were removed due to mobility (lack of osseointegration), erythema of surrounding implant area, vertical bone loss and horizontal ridge deficiency after implant placement. All three implants were lost in the preloading phase. Survival rate of loaded implants was 100%. Table 3 demonstrates the survival rate of implants.

Table 3 - Life table analysis for implants showing the time interval in years, number of implants that existed during these time intervals (N), number of failures (NF), replaced implants (RI), failure rate (Failure%) and survival rate (% S) during this interval, and cumulative survival rate (Cum % S)

Years	N	NF	RI	% Failure	% S	Cum % S
0-1	23	3	0	13.04	86.96	86.96
1-2	18	-	-	0.0	100.0	86.96
2-3	8	-	-	0.0	100.0	86.96
3-4	8	-	-	0.0	100.0	86.96
4-5	8	-	-	0.0	100.0	86.96
5-6	7	-	-	0.0	100.0	86.96
6-7	6	-	-	0.0	100.0	86.96
7-8	5	-	-	0.0	100.0	86.96
>8	5	-	-	0.0	100	86.96

Discussion

In this study, we attempted to assess the feasibility of using EDR data to characterize the dental treatments for SS patients and determine the longevity of dental implants placed. Study results indicate that EDR data could be utilized to characterize the dental treatments in SS patients and assess the effectiveness of these treatments in restoring their oral functions. To the best of our knowledge, this is the first study of dental treatments among SS patients using EDR data. We identified 50% of SS patients received dental restorations using materials such as composite resin and amalgam. Approximately 14% of patients received endodontic treatments to treat infection/disease involving dental pulp. Implants are emerging as a popular alternative to restore lost teeth due to difficulty with tolerating removable denture prosthesis as a result of mucosal dryness. However, we identified only 11 (6.11%) patients who received implant treatment. The high cost and limited coverage of implant treatments under dental insurance in the United States could be reasons for this small number of implant placement. In addition, many of these patients already have high medical expenses due to associated comorbid conditions [38]. We found the average number of dental visits for patients with more than one year of follow-up to be 5.23 visits/year, which is higher than the 4 visits/year reported in a previous study on primary SS patients in the United States [38]. A high number of dental visits with only 41.66% of patients having dental insurance indicated that SS patients incur high dental expenses.

In this study, three implants failed in two patients. The failure rate of approximately 13% on 40 months of follow-up is higher than all the previous studies in SS patients except for the 1999 case review series by Isidor et al. in Denmark, which reported failure rate of 16.7% on 48 months of follow up [9; 11; 19; 22]. Implant failure rate in SS patients is high compared to the 98% success of implants in medically health patients on 10 years of follow-up [22; 25; 39]. Curiously, the

three dental implants that had failed were all during the preloading phase, whether such a trend can be substantiated warrants further investigation. Furthermore, future research in this area is needed to evaluate the impact of different risk factors such as immunosuppressant therapy, associated comorbid conditions and smoking on the implants survival in SS patients.

Several limitations exist within our study. First, SS patients were identified by extracting information from the EDR using the term, "Sjogren". Patients with Sjögren's syndrome whose disease was documented using other lexical variations such as "SJS" would not be identified using our extraction method. Second, Sjogren documentation was based upon patient self-reported data, the reliability of patient self-reported Sjögren's syndrome have not been evaluated. Third, the survival rate of implants within our study was based on 23 implant treatments. The failure rate could be exaggerated due to the limited number of implants.

Future work would be to expand our terminology to include other terms for Sjögren's syndrome. Additionally, we identified the Sjögren's syndrome patient population using information within the EDR. To expand our cohort, we could also use other sources such as medical records for identifying dental patients with Sjögren's syndrome. Furthermore, the accuracy of patient self-reported Sjögren's syndrome needs to be validated with other sources such as medical records.

Conclusion

EDR data could be used for identifying treatments received by SS patients and assessing outcomes. However, further studies are required to evaluate the impact of confounding variables on the outcome of these treatments. Such studies will facilitate developing best practice guidelines to improve oral health among these patients.

Acknowledgements

We acknowledge Mr. Anand Krishnan and Mr. Craig Eberhardt at the Indiana University School of Dentistry for their assistance with data extraction process for this study. We also thank Dr. Mei Song for their valuable comments on the manuscript

References

- [1] R. Abramovicz-Finkelsztain, C.G. Barsottini, and H.F. Marin, Electronic Dental Records System Adoption, *Stud Health Technol Inform* 216 (2015), 17-20.
- [2] K. Albrecht, J. Callhoff, G. Westhoff, T. Dietrich, T. Dörner, and A. Zink, The Prevalence of Dental Implants and Related Factors in Patients with Sjogren Syndrome: Results from a Cohort Study, *J Rheumatol* 43 (2016), 1380-1385.
- [3] J.C. Atkinson and P.C. Fox, Sjogren's syndrome: oral and dental considerations, *J Am Dent Assoc* 124 (1993), 74-76, 78-82,84-76.
- [4] M. Baudet-Pommel, E. Albuissou, J.L. Kemeny, F. Falvard, J.M. Ristori, M.P. Frayse, and B. Sauvezie, Early dental loss in Sjogren's syndrome. Histologic correlates. European Community Study Group on Diagnostic Criteria for Sjogren's Syndrome (EEC COMAC), *Oral Surg Oral Med Oral Pathol* 78 (1994), 181-186.
- [5] P.P. Binon, Thirteen-year follow-up of a mandibular implant- supported fixed complete denture in a patient with Sjogren's syndrome: a clinical report, *J Prosthet Dent* 94 (2005), 409-413.
- [6] S.J. Bowman, G.H. Ibrahim, G. Holmes, J. Hamburger, and J.R. Ainsworth, Estimating the prevalence among Caucasian women of primary Sjogren's syndrome in two general practices in Birmingham, UK, *Scand J Rheumatol* 33 (2004), 39-43.
- [7] A.J. Carr, W.F. Ng, F. Figueiredo, R.I. Macleod, M. Greenwood, and K. Staines, Sjogren's syndrome - an update for dental practitioners, *Br Dent J* 213 (2012), 353-357.

- [8] S. Carsons, A review and update of Sjogren's syndrome: manifestations, diagnosis, and treatment, *Am J Manag Care* 7 (2001), S433-443.
- [9] K. Chochlidakis, C. Ercoli, and S. Elad, Challenges in implant-supported dental treatment in patients with Sjogren's syndrome: A case report and literature review, *Quintessence Int* 47 (2016), 515-524.
- [10] L.B. Christensen, P.E. Petersen, J.J. Thorn, and M. Schiodt, Dental caries and dental health behavior of patients with primary Sjogren syndrome, *Acta Odontol Scand* 59 (2001), 116-120.
- [11] M. de Mendonca Invernici, A. Finger Stadler, G. Vale Nicolau, M.A. Naval Machado, A.A. Soares de Lima, and M. Compagnoni Martins, Management of Sjogren's Syndrome Patient: A Case Report of Prosthetic Rehabilitation with 6-Year Follow-Up, *Case Rep Dent* 2014 (2014), 761251.
- [12] C. Flores-Mir, N.G. Palmer, H.C. Northcott, F. Khurshed, and P.W. Major, Perceptions and attitudes of Canadian dentists toward digital and electronic technologies, *J Can Dent Assoc* 72 (2006), 243.
- [13] R.I. Fox, Sjogren's syndrome, *Lancet* 366 (2005), 321-331. [14] M.F. Furukawa, J. King, V. Patel, C.J. Hsiao, J. Adler- Milstein, and A.K. Jha, Despite substantial progress In her adoption, health information exchange and patient engagement remain low in office settings, *Health Aff (Millwood)* 33 (2014), 1672-1679.
- [15] M. Gandia, E.M. Morales-Espinoza, R.M. Martin-Gonzalez, S. Retamozo, B. Kostov, R. Belenguer-Prieto, D. Buss, M. Caballero, A. Bove, H. Gueitasi, P. Brito-Zeron, A. Siso- Almirall, M.J. Soto-Cardenas, and M. Ramos-Casals, Factors influencing dry mouth in patients with primary Sjogren syndrome: usefulness of the ESSPRI index, *Oral Health Dent Manag* 13 (2014), 402-407.
- [16] A. Hajiabbasi, I. Shenavar Masoooleh, Y. Alizadeh, A.S. Banikarimi, and P. Ghavidel Parsa, Secondary Sjogren's Syndrome in 83 Patients With Rheumatoid Arthritis, *Acta Med Iran* 54 (2016), 448-453.
- [17] C.G. Helmick, D.T. Felson, R.C. Lawrence, S. Gabriel, R. Hirsch, C.K. Kwok, M.H. Liang, H.M. Kremers, M.D. Mayes, P.A. Merkel, S.R. Pillemer, J.D. Reville, and J.H. Stone, Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I, *Arthritis Rheum* 58 (2008), 15-25.
- [18] J. Hu, H. Yu, E. Luo, E. Song, X. Xu, H. Tan, and Y. Wang, Are Chinese dentists ready for the computerization of dentistry? A population investigation of China's metropolises, *J Am Med Inform Assoc* 16 (2009), 409-412.
- [19] F. Isidor, K. Brondum, H.J. Hansen, J. Jensen, and S. Sindet- Pedersen, Outcome of treatment with implant-retained dental prostheses in patients with Sjogren syndrome, *Int J Oral Maxillofac Implants* 14 (1999), 736-743.
- [20] M. Ittah, C. Miceli-Richard, J. Eric Gottenberg, F. Lavie, T. Lazure, N. Ba, J. Sellam, C. Lepajolec, and X. Mariette, B cell- activating factor of the tumor necrosis factor family (BAFF) is expressed under stimulation by interferon in salivary gland epithelial cells in primary Sjogren's syndrome, *Arthritis Res Ther* 8 (2006), R51.
- [21] J.H. John, D. Thomas, and D. Richards, Questionnaire survey on the use of computerisation in dental practices across the Thames Valley Region, *Br Dent J* 195 (2003), 585-590;discussion 579.
- [22] A. Korfage, G.M. Raghoobar, S. Arends, P.M. Meiners, A. Visser, F.G. Kroese, H. Bootsma, and A. Vissink, Dental Implants in Patients with Sjogren's Syndrome, *Clin Implant Dent Relat Res* (2015).
- [23] R.C. Lawrence, D.T. Felson, C.G. Helmick, L.M. Arnold, H. Choi, R.A. Deyo, S. Gabriel, R. Hirsch, M.C. Hochberg, G.G. Hunder, J.M. Jordan, J.N. Katz, H.M. Kremers, and F. Wolfe, Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II, *Arthritis Rheum* 58 (2008), 26-35.
- [24] M. Margaix-Munoz, J.V. Bagan, R. Poveda, Y. Jimenez, and G. Sarrion, Sjogren's syndrome of the oral cavity. Review and update, *Med Oral Patol Oral Cir Bucal* 14 (2009), E325-330.
- [25] H.J. Meijer, G.M. Raghoobar, R.H. Batenburg, and A. Vissink, Mandibular overdentures supported by two Branemark, IMZ or ITI implants: a ten-year prospective randomized study, *J Clin Periodontol* 36 (2009), 799-806.
- [26] J.M. Meijer, P.M. Meiners, J.J. Huddleston Slater, F.K. Spijkervet, C.G. Kallenberg, A. Vissink, and H. Bootsma, Health- related quality of life, employment and disability in patients with Sjogren's syndrome, *Rheumatology (Oxford)* 48 (2009), 1077-1082.
- [27] G. Melenyk, The updated ADA CDT manual, *J Mich Dent Assoc* 90 (2008), 26.
- [28] H. Nakamura, A. Kawakami, and K. Eguchi, Mechanisms of autoantibody production and the relationship between autoantibodies and the clinical manifestations in Sjogren's syndrome, *Transl Res* 148 (2006), 281-288.
- [29] J.J. Napenas and T.S. Rouleau, Oral complications of Sjogren's syndrome, *Oral Maxillofac Surg Clin North Am* 26 (2014), 55-62. [30] A.G. Payne, J.F. Lowmie, and W.J. Van Der Linden, Implant-supported prostheses in patients with Sjogren's syndrome: a clinical report on three patients, *Int J Oral Maxillofac Implants* 12 (1997), 679-685.

Address for correspondence

Zasim Siddiqui, BDS, MS

Indiana University School of Dentistry, 1121 W Michigan Street, Room DS 267, Indianapolis, IN 46202, USA

Email: zasiddiq@iupui.edu Phone: +1-317-274-7130.

Preferred method of contact is through email.

Analysis of Clinical Variations in Asthma Care Documented in Electronic Health Records Between Staff and Resident Physicians

Sunghwan Sohn^a, Chung-Il Wi^b, Young J Juhn^b, Hongfang Liu^a

^aDepartment of Health Sciences Research, Mayo Clinic, Rochester, MN, USA,

^bDepartment of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, USA

Abstract

Clinical documentation using free text to describe a patient's medical status is an essential component of electronic health records (EHRs), and the quality of information in documents plays a critical role in clinical practice and translational research. Physicians are the primary creators of EHRs, but their clinical practices vary substantially, resulting in variations in clinical documentation. These variations can represent a source for potential bias in clinical outcomes and downstream applications using EHRs. Asthma is one example, presenting an inconsistent ascertainment process and criteria. A recent study revealed that resident physicians' knowledge of asthma diagnosis and management is relatively limited. In this study, we examined clinical documentation variations in asthma care between staff and resident physicians using individual words, topics, and asthma-related concepts in EHR clinical narratives. Additionally, we discuss potential biases in building an informatics model and further compare asthma diagnosis and outcomes between two physician groups.

Keywords: Asthma, Documentation, Electronic Health Records

Introduction

Asthma is one of the five most burdensome diseases [1] and the most common chronic childhood illness, affecting approximately 4%-17% of children in the United States [2; 3]. Structured data (e.g., ICD-9 codes) in electronic health records (EHRs) reveal a low sensitivity level to the accurate identification and management of children with asthma. The recent growth of unstructured clinical documents in EHRs and clinical text analytics using machine learning and natural language processing (NLP) have demonstrated great potential in the areas of clinical practice and translational research. These techniques have been successfully applied in asthma [4] and other various clinical applications [5-12].

The clinical documentation of a patient's medical status using free text is an essential component in EHRs and the quality of information accumulated in clinical documents plays a critical role in patient care and clinical research. Physicians are the primary creators of EHRs, but their clinical practices vary substantially [13], which may significantly affect the quality of both structured (e.g., billing codes) and unstructured data (e.g., clinical notes) and cause potential bias in the secondary use of EHR data. Lack of a consistent approach to patient care may also result in higher costs and inevitably makes it necessary to reduce clinical variations at the point of care [14; 15]. For example, asthma has an inconsistent ascertainment process and criteria that cause delays in diagnosis, despite the availability of effective asthma therapies. Further, it has been

reported that junior healthcare providers' (e.g., residents) knowledge of asthma diagnosis and management is in development and needs to be improved through an interactive course [16]. Resident physician documentation often contains inaccurate information in progress notes [17] and as a result, the effect of a point-of-care personal digital assistant has been examined as a means of reducing resident documentation discrepancies [18]. Another study investigates the effect of an EHR template on family medicine resident documentation to improve asthma severity classification and treatment [19].

The quality of clinical documentation is crucial for downstream applications of EHRs to ensure the strong performance of informatics systems developed based on them. However, there are gaps between experienced (i.e., staff) and inexperienced (i.e., residents) healthcare providers in clinical documentation, and this can result in unintended consequences. To build accurate informatics systems, clinical variations between two provider groups need to be identified and addressed. Inappropriate or missing representations of a patient's medical status as a result of inexperienced healthcare providers may diminish preventive opportunities and lead to inefficient healthcare delivery. This challenges physicians in their efforts to improve the value of healthcare delivery.

To address this problem, it is important to identify major variations in clinical documentation and understand the nature of those variations. Previously, researchers have investigated differences in laboratory tests in a physician's practice to identify overlooked information regarding physician behavior [13]. The association between residency training and clinical practice [20; 21] was studied to identify variations in terms of clinical education. Regional variances among healthcare providers were also explored in diagnostic practices measured in Medicare claims [22]. Although it has been observed that the recent trend of novel efforts to promote rational use may cause variations in physician practice [23], the nature of such variations are not well understood. These previous studies utilized structured data and educational backgrounds to analyze physician behavior rather than the rich source of unstructured data in EHRs (i.e., clinical narratives).

In this study, we assessed variations in clinical documentation, especially clinical narratives, between staff and resident physicians in asthma care using individual words, topics, and asthma-related concepts. Further, we compared actual asthma diagnoses and outcomes between two physician groups and discussed the potential impacts of clinical document variations on quality of care and an informatics model.

Methods

We examined basic statistics and similarities in clinical documents between staff and resident physicians to identify clinical

documentation variations. Asthma diagnoses and outcomes statistics were then compared to analyze gaps in clinical practice. This study was approved by the Mayo Clinic Institutional Review Board. Detailed methods are described in the following subsections.

Patient Cohort

Our study subjects are children who received medical care from Mayo Clinic pediatric practice, Rochester, Minnesota. As of 2015 September, 19,064 children were cared for by Mayo Clinic pediatric practice. Of these, 9,277 children were included for this study who participated in an ongoing asthma intervention study.

To examine document statistics and similarities we used the all patient cohort ($n=9,277$) and associated clinical notes. For the comparison of asthma diagnoses and outcomes, we selected patients from 2013 to 2015 who had clinical visits with either staff or resident physicians in Community Pediatric & Adolescent Medicine (CPAM) based on the 2015 physician list ($n=4,102$).

Clinical Document Statistics and Similarities

Basic statistics on clinical documents (i.e., total number of clinical documents/tokens, frequency of asthma-related concepts, note types, and sections in clinical notes) were compared between staff and residents. Staff and resident clinical document similarities were also compared in terms of word-level, topic-level, and asthma-related concepts. We preprocessed the documents (i.e., tokenization, removing stop words, and stemming) and created a vector space model for each case to compare similarity. The similarity was measured by cosine similarity since it is useful for comparing how two documents or corpora are likely to be in terms of their subject matter [24].

The clinical documents were represented by tf-idf (term frequency-inverse document frequency) and latent topic-based vector for computing word-level and topic-level similarity, respectively. The tf-idf for the term t is defined by summation of $tf(t)idf(t)$ for all documents in the corpus divided by total number of documents in the corpus, i.e., $\sum_i tf(t)idf_i(t)/N$.

In order to compare clinical documents in the latent topic space, we employed a topic modeling method, Latent Dirichlet Allocation (LDA) [25; 26], to generate the document distributions in the latent topic space, i.e., $p(z_k|d_i)$. In the latent topic-based vector representation, the topic z_k for the corpus C is defined as:

$$p(z_k|C) = \sum_{d_i \in C} p(z_k|d_i, C)p(d_i|C) = \sum_{d_i \in C} \frac{p(z_k|d_i)}{N}$$

To compute similarity of asthma-related concepts we extracted concepts used in the predetermined asthma criteria (PAC) (see Table 1) and represented them in a vector space model using the same notion of tf-idf representation.

Asthma-related Concept Extraction

We used the MedTaggerIE module [27] in MedTagger to extract asthma-related events and episodes. MedTagger is the open-source pipeline developed by Mayo Clinic that contains a suite of programs including three major components: indexing based on dictionaries, information extraction based on patterns, and machine learning-based named entity recognition [27; 28]. Asthma-related concepts (Table 1) were compiled from PAC [29] that were originally developed by Yunginger

et al.[29] and have been used extensively in research for asthma epidemiology. Each concept consists of a set of keywords.

Table 1. Asthma-related concepts used in predetermined asthma criteria.

Cough, Asthma, Infantile Eczema, Wheeze, Night-time disturbance (nocturnal cough/wheezing), Hay fever, Dyspnea, Bronchiolitis, Pulmonary Test, Bronchodilator, Bronchospasm, Methacholine Test, Positive Skin Test, COPD, Nasal Polyps
--

Asthma Diagnoses and Outcomes

We compared asthma diagnosis statistics based on both physician-diagnosis and PAC between staff and residents. Physician-diagnosed asthma was determined by examining the diagnosis section in clinical notes, and PAC-based asthma diagnosis was determined by applying our NLP-based PAC system [4]. Further we compared outcomes of asthma care, such as ER visits for asthma, hospitalization for asthma, and asthma exacerbation (i.e., oral steroid use for asthma).

Results

Clinical Document Statistics

The basic descriptive clinical document statistics of staff vs. resident physicians are shown in Table 2. The list of providers (i.e., staff or residents) was based on the year 2015. Staff have a higher number of patients and documents because they tend to stay longer than residents at the clinic. Interestingly, residents have a much higher median number of tokens (words) per document than staff. Both staff and residents have a similar proportion of documents that contain asthma-related concepts and median number of asthma-related concepts per document.

Figure 1 shows the distribution of asthma-related concepts between staff and residents. The concept *cough* appears most dominant in both providers' notes, but staff documented it more frequently than residents. The next most frequent concepts for staff were *asthma*, *eczema*, and *wheezing*, and for residents, *asthma*, *dyspnea*, *eczema*, and *wheezing* (*asthma* and *dyspnea* appear in very similar patterns). Residents' documents have a much higher frequency for *dyspnea* than staff. It should especially be noted that *asthma* appeared less frequently in resident documentation, even though the other concepts appeared in similar proportion, including *wheezing*, which is the most important concept in determining a patient's asthma status.

Figure 2 shows the distribution of note types that contain asthma-related concepts. Both staff and residents recorded asthma-related concepts dominantly in limited exam (LE) and multi-system evaluation (ME) notes. Residents also recorded these concepts in progress (PRG) notes and also used admission (ADM), observation (AOB), and consultant (CON) notes, while staff rarely used these note types.

The distribution of sections that contain asthma-related concepts between staff and residents is shown in Figure 3. Staff used history of present illness (HPI) most frequently, followed by impression/report/plan (IRP), but the order is reversed for residents. Residents also used the subjective section, which was rarely used by staff.

Clinical Document Similarities

Similarities between staff and resident clinical documents were examined in terms of word-level, topic-level, and asthma-related concepts (Table 3). Word-level similarity was not high, but topics and asthma-related concepts between two provider groups were very similar.

Asthma Diagnosis and Outcome

Table 4 contains statistics pertaining to asthma diagnoses and outcomes for patients whose CPAM physicians have been

either staff or residents for the past three years (2013 to 2015). Staff diagnosed asthma more than twice as often as residents in both physician-diagnosed (10.4% and 4.5%, respectively) and PAC-based asthma (10.5% and 4.6%, respectively) when compared based on the proportion of their patients. For each of three asthma outcomes (i.e., ER visit for asthma, hospitalization for asthma, and asthma exacerbation), staff demonstrated better quality of care (i.e., lower ratio of incidents) than residents and they were statistically significant (95% CI) when considering all three outcomes as a whole.

Table 2. Basic statistics of clinical notes written by staff and residents for pediatric patients

	Staff (n=21)	Resident Physicians (n=56)
Date range	1998 - 2015	2011 - 2015
No. patients seen by staff or residents	8,434	4,252
Total no. documents	97,443	16,268
No. documents/year, median (IQR)	4,888 (5,491)	2,120 (5,238)
No. tokens/document, median (IQR)	274.0 (392)	530.5 (655)
No. documents with asthma-related concepts (%)	22,504 (23.1)	4,301 (26.4)
No. asthma-related concepts/document, median (IQR)	2 (4)	2 (4)

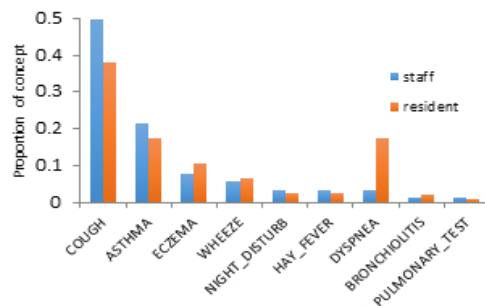


Figure 1. Asthma-related concept distribution (proportion $\geq 1\%$).

Discussion

Clinical documentation between staff and resident physicians in asthma care has been examined in terms of document contents, asthma diagnosis, and asthma outcomes. Basic document statistics (i.e., the number of documents, tokens, asthma-related concepts, and note types) differed in number, and the individual words used in documents were not so similar (word-level similarity = 0.718). However, concepts used in their documents were very similar (i.e., topic similarity = 0.991; asthma-related concept similarity = 0.964), which may reflect the fact that both staff and residents share common terms to describe asthma episodes and events.

For asthma-related concepts in clinical documents (Figure 1), staff and residents demonstrated similar distribution except for the concept of *dyspnea* (consists of keywords: *dyspnea*, *shortness of breath*, *breathing difficulty*, *respiratory distress*, etc.). Interestingly, residents recorded a much larger portion of *dyspnea* (17.3%) than staff (3.2%), which is close to the proportion of residents' recorded *asthma* concepts (17.5%). Staff

Table 3. Similarities of clinical documents between staff and resident physicians

	Cosine similarity
Word	0.718
Topic	0.991
Asthma-related concept	0.964

and residents showed different usage of note types that contained asthma-related concepts (Figure 2), although the top two frequently used note types were the same (limited exam, multi-system evaluation). Over 90% of asthma-related concepts were found in three note types for staff (limited exam, multi-system evaluation, subsequent visit), while residents used additional note types (progress, admission), up to 90% of asthma-related concepts. We also investigated overall note type distribution regardless of asthma-related concepts and observed that progress notes were the second most frequently used notes for residents (23%) after multi-system evaluation (23%). Staff rarely used admission and observation notes, which may be one reason for note type discrepancies that contained asthma-related concepts. The overall trend of using sections for asthma-related concepts (Figure 3) was similar between two physician groups, but interestingly, residents used impression/report/plan sections most frequently, as compared to history of present illness, which was most commonly used among staff. Additionally, residents did not use diagnosis sections as often as staff. This may reflect the fact that residents are less likely confirmative in their asthma-related observations with regard to patient medical status.

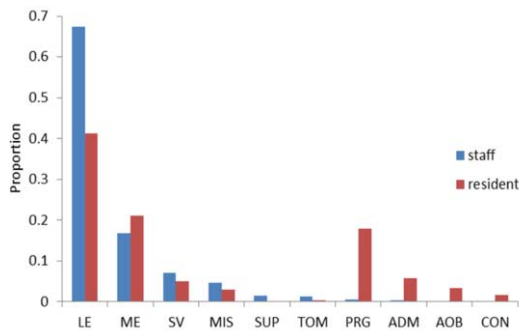


Figure 2. Distribution of note types that contains asthma-related concepts (proportion $\geq 1\%$). LE: limited exam, ME: multi-system evaluation, SV: subsequent visit, MIS: miscellaneous, SUP: supervisory, TOM: test-oriented miscellaneous, PRG: progress, ADM: admission, AOB: observation, CON: consultation

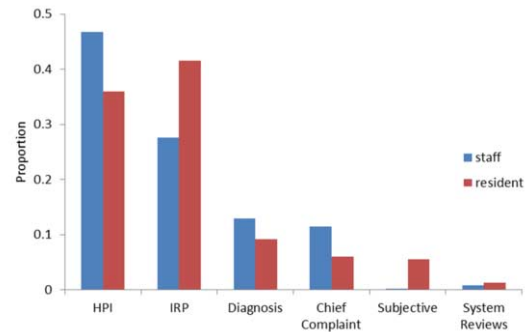


Figure 3. Distribution of sections that contains asthma-related concepts (proportion $\geq 1\%$). HPI: history of present illness, IRP: impression/report/plan

Table 4. Asthma diagnoses and outcomes for staff and resident patients (2013 to 2015)

		Staff	Resident
Basic statistics	No. patients	3,304	798
	Median No. documents/patient (IQR)	3 (5)	2 (2)
Asthma diagnosis	No. patients of physician diagnosed asthma (%)	342 (10.4)	36 (4.5)
	No. patients of PAC asthma (%)	348 (10.5)	37 (4.6)
Outcome[†]	ER visit for asthma, ratio	0.067	0.306
	Hospitalization for asthma, ratio	0.012	0.083
	Asthma exacerbation (oral steroid use), ratio	0.211	0.333
	Any of above three outcomes (p-val = 0.034)	0.234	0.417

[†] No. patients with a given condition/No. patients with CPAM physician-diagnosed asthma

There were notable gaps in clinical documentation between staff and resident physicians in asthma care. Our study shows that residents were less likely to diagnose asthma than staff even though the distribution of other asthma-related concepts appeared to be similar. Subsequently, asthma outcomes (ER visits, hospitalization, and oral steroid use) among patients of residents were relatively poor when compared with those of staff.

The recent growth of EHRs provides potential for comprehensive tracking of patient history, which allows evidence-based medicine to address treatment uncertainty and reduce clinical variances. However, EHRs may contain hidden biases and noise factors unrelated to disease burden [30; 31]. Without careful analysis of such biases, an informatics model built on such EHRs may not perform as expected. Along this line, differences in clinical documentation between staff and residents observed in our study must also be considered in order to ensure the accuracy of informatics models.

The similarity measure of asthma-related concepts used in our study depends on the type of concepts in the entire documents of each physician group, not necessarily individual numbers of each concept per physician. The high similarity reflects that both staff and residents use similar concepts, and therefore residents may not be required to undergo additional education to ensure proper use of additional terms or definitions. However, it does not show the actual presence/absence of individual concepts representing a patient's medical status per physician. Further assessment of detailed presence/absence concepts of residents' documentation representing patient medical

status and divergence analysis compared to staff could be a strong potential resource, not only in the education of residents, but also in an effort to improve clinical documentation so as to reduce clinical variations.

Limitations of this study include lack of analyzing effects of potential confounding factors (e.g., patient's demographic, environment, risk factors, and healthcare utilization) and the use of a relatively short period of time (three years) to analyze and compare asthma diagnoses and outcomes. We plan to consider these factors in the future and expand divergence analysis in asthma care at the individual resident level.

Conclusions

Clinical documentation variations were observed in asthma care between staff and resident physicians when utilizing note types, sections, and ascertaining asthma-related concepts. Additionally, residents were less likely to diagnose asthma than staff and were responsible for relatively poor outcomes in asthma care because a delay in asthma diagnosis may prevent asthmatic children from access to therapeutic and preventive interventions. The variations identified by our analysis may represent a potential bias, not only for clinical outcomes, but also for downstream applications using EHRs, and this needs to be considered in order to better interpret clinical research outcomes and the performance of informatics models for clinical decision support.

Acknowledgements

This work was made possible by NIGMS R01GM102282, NLM R01LM11934, NIBIB R01EB19403, and NHLBI R01HL126667.

References

- [1] M.W. Stanton and M. Rutherford, *The high concentration of US health care expenditures*, Agency for Healthcare Research and Quality Washington, DC, 2006.
- [2] Centers for Disease Control and Prevention, Vital signs: asthma prevalence, disease characteristics, and self-management education: United States, 2001–2009, *MMWR. Morbidity and mortality weekly report* 60 (2011), 547.
- [3] D.L. Blackwell, J.W. Lucas, and T.C. Clarke, Summary health statistics for US adults: national health interview survey, 2012, *Vital and health statistics. Series 10, Data from the National Health Survey* (2014), 1–161.
- [4] S.T. Wu, S. Sohn, K. Ravikumar, K. Waghlikar, S.R. Jonnalagadda, H. Liu, and Y.J. Juhn, Automated chart review for asthma cohort identification using natural language processing: an exploratory study, *Annals of Allergy, Asthma & Immunology* 111 (2013), 364–369.
- [5] S. Sohn, C. Clark, S. Halgrim, S. Murphy, C. Chute, and H. Liu, MedXN: an Open Source Medication Extraction and Normalization Tool for Clinical Text, *Journal of the American Medical Informatics Association* 21 (2014), 858–865.
- [6] S. Sohn, J.-P.A. Kocher, C.G. Chute, and G.K. Savova, Drug side effect extraction from clinical narratives of psychiatry and psychology patients, *J Am Med Inform Assoc* 18 (2011), 144–149.
- [7] S. Sohn and G.K. Savova, Mayo Clinic Smoking Status Classification System: Extensions and Improvements, in: *AMIA Annual Symposium*, 619–623, San Francisco, CA, 2009.
- [8] S. Sohn, Z. Ye, H. Liu, C. Chute, and I. Kullo, Identifying Abdominal Aortic Aneurysm Cases and Controls using Natural Language Processing of Radiology Reports., in: *AMIA Summits Transl Sci Proc*, San Francisco, 2013.
- [9] S. Sohn, M. Torii, D. Li, K. Waghlikar, S. Wu, and H. Liu, A Hybrid Approach to Sentiment Sentence Classification in Suicide Notes, *Biomedical Informatics Insights* 2012, 43–50.
- [10] D. Demner-Fushman, W. Chapman, and C. McDonald, What can natural language processing do for clinical decision support?, *Journal of Biomedical Informatics* 42 (2009), 760–772.
- [11] I.J. Kullo, K. Ding, H. Jouni, C.Y. Smith, and C.G. Chute, A genome-wide association study of red blood cell traits using the electronic medical record, *PLoS One* 5 (2010), e13011.
- [12] S.V.S. Pakhomov, J.D. Buntrock, and C.G. Chute, Automating the assignment of diagnosis codes to patient encounters using example-based and machine learning techniques, *Journal of the American Medical Informatics Association* 13 (2006), 516–525.
- [13] G.M. Weber and I.S. Kohane, Extracting physician group intelligence from electronic health records to support evidence based medicine, *PLoS One* 8 (2013), e64933.
- [14] J.D. Schuur, C.W. Baugh, E.P. Hess, J.A. Hilton, J.M. Pines, and B.R. Asplin, Critical Pathways for Post-Emergency Outpatient Diagnosis and Treatment: Tools to Improve the Value of Emergency Care, *Academic Emergency Medicine* 18 (2011), e52–e63.
- [15] J.E. Wennberg, Unwarranted variations in healthcare delivery: implications for academic medical centres, *British Medical Journal* 325 (2002), 961.
- [16] A.R. Hemnes, A. Bertram, and S.D. Sisson, Impact of medical residency on knowledge of asthma, *Journal of Asthma* 46 (2009), 36–40.
- [17] A.E. Carroll, P. Tarczy-Hornoch, E. O'Reilly, and D.A. Christakis, Resident documentation discrepancies in a neonatal intensive care unit, *Pediatrics* 111 (2003), 976–980.
- [18] A.E. Carroll, P. Tarczy-Hornoch, E. O'Reilly, and D.A. Christakis, The effect of point-of-care personal digital assistant use on resident documentation discrepancies, *Pediatrics* 113 (2004), 450–454.
- [19] A.M. Davis, M. Cannon, A.Z. Ables, and H. Bendyk, Using the electronic medical record to improve asthma severity documentation and treatment among family medicine residents, *Family Medicine* 42 (2010), 334.
- [20] B.E. Sirovich, R.S. Lipner, M. Johnston, and E.S. Holmboe, The association between residency training and internists' ability to practice conservatively, *JAMA Internal Medicine* 174 (2014), 1640–1648.
- [21] E.F. Haponik, D.L. Bowton, R. Chin, N.E. Adair, M.G. Lykens, P.T. Alford, and D.A. Bass, Pulmonary section development influences general medical house officer interests and ABIM certifying examination performance, *CHEST Journal* 110 (1996), 533–538.
- [22] Y. Song, J. Skinner, J. Bynum, J. Sutherland, J.E. Wennberg, and E.S. Fisher, Regional variations in diagnostic practices, *New England Journal of Medicine* 363 (2010), 45–53.
- [23] C.K. Cassel and J.A. Guest, Choosing wisely: helping physicians and patients make smart decisions about their care, *JAMA* 307 (2012), 1801–1802.
- [24] A. Singhal, Modern information retrieval: A brief overview, *IEEE Data Eng. Bull.* 24 (2001), 35–43.
- [25] D.M. Blei, A.Y. Ng, and M.I. Jordan, Latent dirichlet allocation, *The Journal of Machine Learning Research* 3 (2003), 993–1022.
- [26] Y. Wang, J.S. Lee, and I.C. Choi, Indexing by latent dirichlet allocation and an ensemble model, *Journal of the Association for Information Science and Technology* (2015).
- [27] H. Liu, S. Bielinski, S. Sohn, S. Murphy, K. Waghlikar, S. Jonnalagadda, R. KE, S. Wu, I. Kullo, and C. Chute, An information extraction framework for cohort identification using electronic health records, in: *AMIA Summits Transl Sci Proc*, San Francisco, CA, 2013, pp. 149–153.
- [28] M. Torii, K. Waghlikar, and H. Liu, Using machine learning for concept extraction on clinical documents from multiple data sources, *Journal of the American Medical Informatics Association* 18 (2011), 580–587.
- [29] J.W. Yunginger, C.E. Reed, E.J. O'Connell, L.J. Melton III, W.M. O'Fallon, and M.D. Silverstein, A community-based study of the epidemiology of asthma: incidence rates, 1964–1983, *American Review of Respiratory Disease* 146 (1992), 888–894.
- [30] S. Dey, J. Cooner, C.W. Delaney, J. Fakhoury, V. Kumar, G. Simon, M. Steinbach, J. Weed, and B.L. Westra, Mining Patterns Associated With Mobility Outcomes in Home Healthcare, *Nursing Research* 64 (2015), 235–245.
- [31] P. Yadav, M. Steinbach, V. Kumar, and G. Simon, Mining Electronic Health Records (EHR): A Survey, (2015).

Address for correspondence

Sunghwan Sohn; Mayo Clinic, 200 First St SW, Rochester, MN 55905; Sohn.Sunghwan@mayo.edu

Medical Effect of Venous Thromboembolism Prophylaxis Systems and Common Input Categories: Preliminary Findings from a Systematic Review

Julian Varghese^a, Maren Kleine^b, Sophia Isabella Gessner^a, Sarah Sandmann^a, Martin Dugas^a

^a Institute of Medical Informatics, University of Münster, Münster, Germany

^b School of Public Health, Bielefeld University, Bielefeld, Germany

Abstract

Computerized Clinical Decision Support Systems (CDSS) are implemented in hospitals to improve prevention of Venous Thromboembolisms (VTE). A physician-driven review was conducted to assess extent of patient outcome effects of recently published CDSS studies. To facilitate future re-implementations within existing hospital information systems, input variables of included systems were extracted, standardized and annotated with semantic codes. Item category coverages of the different systems were then compared. 73% of studies showed positive medical effect. Of these, 53% showed strong positive medical effect by reducing incidence of deep vein thromboses and pulmonary embolisms. Outcome-improving systems tend to cover more item categories. A broad set of clinically relevant input variables should be taken into account or reused from the electronic health record if considering CDSS implementation. Input data models are provided for download in different standardized formats. Site-specific organizational factors that determine how systems are introduced, implemented and tested are also crucial for success.

Keywords:

Clinical Decision Support Systems, Venous Thromboembolism, Common Data Elements

Introduction

Computerized Clinical Decision Support Systems (CDSS) are targeted to improve clinical decision making using individual patient characteristics to generate health-related recommendations with an aim to improve health care quality, clinical processes or cost-efficiency [1]. This work reports on preliminary findings of an ongoing systematic review on CDSS in inpatient settings but focusses on one disease entity where CDSS have been frequently applied: Venous thromboembolisms (VTE). VTE are responsible for a great number of complications in hospitalized patients. They are also one of the most common preventable causes of in-hospital deaths [2]. VTE represent a disease entity where CDSS could improve patient care at sites where guideline-adherent prophylaxis of VTE is under-utilized [3]. Comparing medical effectiveness of different CDSS studies leads to an analysis of different patient outcome measurements and effect sizes; and all of them weighing differently on the patient's overall medical condition. For instance, patient outcome improvement in VTE prophylaxis is linked to many different patient outcome measurements such as number of events of deep vein thromboses or lung embolisms

and adverse reactions to prophylaxis such as hemorrhages [4]. To our knowledge, there is no systematic review on CDSS that assigned full patient outcome summaries to different levels of clinical relevance. We deal with this issue using a new generic method, which intends to rate medical effect of CDSS interventions regardless of their targeted disease entity. The method is used by two physicians for medical evaluation and two biostatisticians for study quality assessment to analyze both clinical impact and risk of bias based on included CDSS studies.

There are regulatory and technical challenges when re-implementing CDSS from a study context into daily routine of hospital information systems. CDSS can also pose patient risk because of health provider's dependency on its proper functioning. Regulatory frameworks approving CDSS software as a medical device need to be considered in the United States [5], the European Union [6] or elsewhere. Regarding technical challenges, necessary input data elements of the system might not be provided in an adequate way before a CDSS can even start to operate on any patient-related data. This leads to issues such as cumbersome manual data re-entry or tiresome mapping of available patient data to the input database of the CDSS [7]. The latter issue is also a result of lacking semantic interoperability in hospital information systems [8]. Data forwarding from primary data capturing systems to a desired database can result in data acquisition time delays or makes transfer of complex patient data challenging [9]. At the level of principal knowledge base of a CDSS the knowledge acquisition bottleneck describes the difficulty of acquiring domain-specific knowledge by experts [10]. CDSS solutions to improve guideline adherence might be hampered due to poor machine-readability of unstructured guideline text. Therefore, input variables and inference need to be clearly defined to facilitate CDSS implementations.

The main objectives of our review are summarized in the following:

1. Medical evaluation of CDSS impact on patient outcomes in inpatient settings and taking into account risk of bias assessment by using a standardized approach.
2. Identification and analysis of different item categories that were used by the systems. By doing this, we also approach to test whether outcome-improving CDSS studies implemented different or more item categories for input than outcome-neutral CDSS.
3. Generate a set of standardized electronic forms including metadata definitions that represent the input of all included CDSS studies.

Methods

The *a priori* design of the ongoing systematic review has been registered on PROSPERO register for systematic reviews [11]. This work focusses on VTE as a specific disease entity and elaborates on common input variables of CDSS that were applied as study interventions with patient outcome measurements. Two standardized tools have been used to provide high quality for systematic reviews: The 27-item checklist “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) [12] and the 11-item checklist of “Assessment of Multiple Systematic Reviews” (AMSTAR) [13]. Statement of checklist fulfillment is provided on our online supplement [14].

Search Strategy

Full text studies on CDSS in English were searched in MEDLINE, Cochrane Trials and Cochrane Reviews from January 2005 to April 2016; see table 1. Article references of all included articles and reviews were taken into consideration for further inclusion. A broad set of search terms was used based on recent systematic reviews [1, 15] on Health Information Systems and focusing on terms relevant for CDSS and outcome assessments. Two search phases were required for entire screening and analysis. The first started in July 2015 (articles screened from 01/2005 till 07/2015) and a second updated search started in April 2016 (articles screened from 07/2015-04/2016). Details of CDSS definition and exact search string are provided on the supplement [14]. CDSS studies suggested by expert opinions collected on MEDINFO Conference 2015 and Medical Informatics Europe Conference 2016 [16] were also taken into consideration.

Table 1– Selection criteria for identifying CDSS studies

Inclusion	Exclusion
<ul style="list-style-type: none"> • CDSS used as main study intervention component aiming to improve VTE prophylaxis • Applied in inpatient hospital or intensive care setting • Used within existing clinical routine workflow operating on real patient data • Evaluation of at least one patient outcome measure with existing control group • Control group has to represent usual care with no components of CDSS being involved. 	<ul style="list-style-type: none"> • Study participants were test patients or CDSS was used solely by medical students • Same CDSS for same disease domain by same site (the latest study was included) • Studies only evaluating guideline adherence or compliance without an evaluation of at least one patient outcome • No English full text article available

Study Quality Evaluation of Included Studies

Study design characteristics such as outcome variables (primary and secondary), study period, principal study design, number of study arms, healthcare setting (e.g. single or multiple hospitals involved), study country, patient sample size, sample size calculation, statistical methods for baseline analysis and confounder adjusting were extracted by two independent reviewers wherever reported. Principal study designs were: 1) randomized controlled, 2) non-randomized interventional or pre/post design with at least one prospective study arm and 3) purely retrospective studies as retrospective cohorts or case-control studies. Randomization is distinguished by unit of randomization: 1) Patient level, 2) Health Care Professional, and 3) Site or ward. Randomization quality of CDSS studies improves in this order due to decreased probability of contamination

[17]. Based on availability of reported characteristics, two biostatisticians used a standardized approach [18] to independently evaluate overall risk of bias (low, medium, high) for each study. Due to the heterogeneity of different principal study designs a meta-analysis for outcome analysis was not considered. Ratings of both reviewers were compared later, in case of differences, agreement was reached by consensus.

Coverage of Input Item Categories

For every included study, one health-IT expert was engaged to identify input data elements, which were required for the CDSS algorithm. A three point rating scale was used to assess reporting quality of such data elements (*poor*: Most or all input variables are not described, reproduction of the required input is not possible; *questionable*: Partially vague descriptions, reproduction of required input is uncertain; *sufficient*: Variables are reported sufficiently enough to reproduce the required input). Assessment was done by reviewer’s judgement on completeness and comprehension of reported data elements to reproduce the CDSS’ input that is necessary for the execution of the algorithm. To facilitate future re-implementations, for each study, reported input variables, associated data types and value-sets were identified and standardized within Operational Data Model (ODM) [19] and manually annotated with semantic codes from the Unified Medical Language System (UMLS) by a physician and uploaded to our medical data repository [20]. We compare item category coverage in outcome-improving vs. outcome-neutral CDSS studies.

After standardization of different input variables to ODM forms, CDEGenerator [21] was used to automatically identify common medical concepts. A medical concept was defined as common if it occurred at least in two of the included CDSS studies as an input variable. Based on clinical similarities, the common medical concepts were assigned to basic routine documentation categories. Thus, an overview of common input variable categories and their use in patient outcome-improving CDSS studies vs. outcome-neutral studies was established.

Medical Evaluation

Reported outcome measurements and results were rated independently by two physicians. For each study, they were instructed to assign all patient outcome effects to a level of clinical relevance. Clinical relevance represents impact on the patient’s overall medical condition using all reported outcome findings, taking into account both statistical significance and actual effect sizes. Similar to American Society of Anesthesiologists’ (ASA) classification [22] used to rate the patient’s preoperative health, the method both physicians used assessed relatedness to life-threatening conditions or mortality as a key indicator. By definition, studies significantly reducing mortality received the highest rating. The following point scale was used, which will be referred to as the medical effect score:

- 5: Mortality was reduced.
- 4: Strong positive effect: No effect on mortality was measured, but clinical events (e.g., adverse events or forms of morbidity) with immediate life threatening potential were reduced.
- 3: Medium positive effect: Clinical events with no immediate life threatening potential were reduced. Patients suffering from those events would require non-urgent treatment.
- 2: Light positive effect: Clinical events with no immediate life threatening potential were reduced. Patients suffering from those events do not necessarily require treatment.
- 1: No effect on medical condition was measured, or outcome findings were mixed (positive and negative outcome results).

A potential benefit for the patient's medical condition is unclear or not expected. If a study had a negative outcome summary it was rated as "negative". All ratings except for 1) required statistical significance. Weighted Kappa statistics were calculated to measure inter-rater reliability before consensus was reached. Details of calculation are provided in the supplement [14].

Results

Search Selection

Based on a set of 70 eligible CDSS studies from our ongoing systematic review [11], 15 studies focused on VTE prophylaxis; see figure 1. The full list of included articles and the rest of all excluded articles can be found in the supplement [14].

Study Characteristics

Non-randomized studies with a pre-post analysis or cohort studies with prospective data collection as a principal study design were most common (47%, $n=7$), followed by purely retrospective ones (33%, $n=5$) and then by randomized controlled studies (20%, $n=3$). 12 (80%) studies implemented CDSS in single hospitals while three managed to run CDSS in multiple hospitals. None of the studies reported on approval for medical device. Further details of study characteristics such as sample size, study duration, adjusting for confounders, study settings, study countries, and numerical outcome results are provided in the supplement [14].

Medical Evaluation

Figure 2 illustrates proportions of medical effect score values and risk of bias. Based on the two physicians review (weighted Kappa: 0.77, $p<0.001$), three out of 15 studies were evaluated to have a medium positive effect (score=3) due to a significant reduction of deep vein thromboses (DVT). Eight studies were evaluated to have a strong medical effect (score=4) as they additionally showed significant reduction of pulmonary embolisms. None of the studies were associated with high risk of bias or negative patient outcomes, e.g. due to an increased number of adverse events.

Technical Characteristics

Eleven studies out of 15 included studies (73%) provided sufficient information on input variables of the systems. Three studies provided questionable reporting quality and one study provided no information on input at all. Reasons for questionable reporting quality were vague descriptions on input variables, e.g., notations like "the algorithm is based on guideline XY" and XY lacks details for proper formalization.

Based on reported information, all systems were rule-based. Fourteen studies (93%) implemented CDSS as alerting tools and provided therapeutic recommendations specifically to the patient context. Only one system output alerts without further clinical advice and showed no impact on patient outcomes [23].

Fourteen out of 15 studies provided information on input data elements. Based on UMLS coding, CDEGenerator identified 34 common medical concepts. Based on clinical similarities of those concepts, routine documentation categories were assigned.

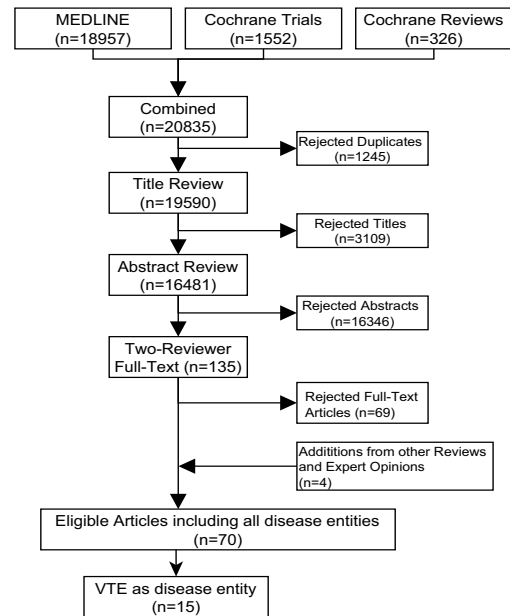


Figure 1– Search strategy of this work is based on our registered ongoing systematic review but with a focus on VTE as the disease entity

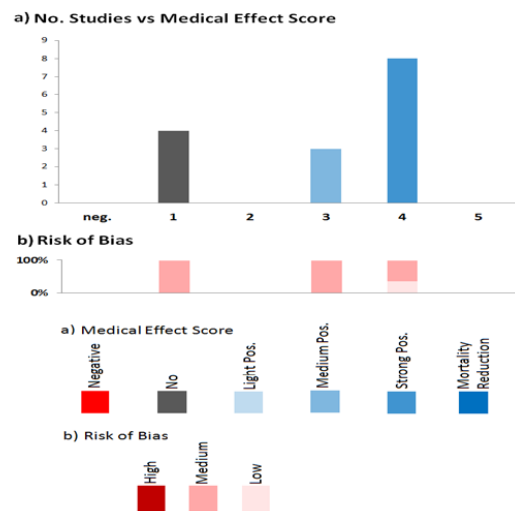


Figure 2– Evaluation of medical effect (a) and risk of bias (b). E.g., eight studies showed a strong positive medical effect of which approximately one third (3/8) were rated with low risk of bias.

Medical history was the most covered category dealing with different disease entities and was divided into further subgroups to represent the different disease areas; see table 2. Comparison of outcome-improving (medical effect score >1) vs. outcome-neutral (medical effect score=1) studies indicated that outcome-improving studies CDSS made use of more item categories overall. For each item category – except for *Medical History of Vascular diseases* – there are relatively more outcome-improving studies that covered the category than the outcome-neutral studies did. The biggest coverage

difference occurred in *Demographics* (90% vs. 50%) where *age* was required as input. Due to the low number of studies in both groups, statistical significance could not be detected (Fisher-Exact Test) within a single category. The full list of medical concepts within the categories is available in the supplement [14]. All input data elements for each study were standardized as ODM forms and annotated with UMLS codes. For technical reuse in diverse information systems, ODM input forms are available in different formats on our medical forms repository [24]. Figure 3 exemplifies an extract of a standardized form that represents input of the CDSS in the study by Mitchell [25], which covered most of all item categories.

Figure 3— Extract of an electronic input form. Detailed information as variable definitions and semantic UMLS codes are available within various formats for reuse in different medical information systems.

Discussion

Regarding study quality, majority of the studies were evaluated with *medium risk of bias*, which means study results are deemed to represent true treatment effect. The study is susceptible to some bias but the problems are not sufficient to invalidate the results [18]. The main reasons for not obtaining evaluation of *low risk* were failures to provide large patient sample sizes, sufficient adjustment for confounders and proper randomization. It has to be noted that randomization – other than drug studies – should not be applied on the patient level but on the level of health providers or hospital wards due to the mentioned contamination issues. Future CDSS studies should address those aspects to provide high study quality. Though all studies received study approval none of them reported on medical device approval. For dissemination of CDSS from a study context into daily routine practice, approval as a medical device is inevitable for systems that had shown to effect clinical processes or patient outcomes [5,6].

Limitations

Due to lack of existing standardized instruments for clinical CDSS evaluation we provided our own non-standardized evaluation method. It proved to be an easy-to-use protocol to link different outcome measurements even from different disease entities to a uniform level of clinical relevance based on physicians' perception of mortality relatedness. In this work, weighted Kappa analysis indicated high inter-physician agreement before consensus was reached. Our criterion to identify CDSS effect solely based on patient outcome meas-

urements is quite strict, since many studies do not manage to show significant changes due to limited patient sample sizes and observation time. Outcome factors such as quality of life or mental health may contribute substantially to patient's overall health, but these were not reported within included CDSS studies. At the level of input variables, we could observe that outcome-improving systems covered more item categories, which probably is an informational advantage to predict individual patient risk with high sensitivity and specificity. But of course, as with all CDSS studies, site-specific clinical, contextual or organizational factors [26] might have been crucial for success as well as further technical characteristics regarding technical interoperability with the hospital information system.

As with all reviews, selective outcome reporting or publication bias is a given since positive outcome effects are more likely to be published than ones that were non-significant or negative [27].

Table 2—Item Categories used in outcome-neutral(–) vs. outcome-improving(+) CDSS studies. MH=Medical History

Category	Outcome– (n=4)	Outcome+ (n=10)
MH: Thromboembolisms	50%	80%
MH: Bleeding	0%	50%
MH: Malignant Neoplasms	50%	80%
MH: Cardiovascular Diseases	25%	60%
MH: Coagulation Disorders	25%	40%
MH: Pregnancy + Births	25%	30%
MH: Lung Diseases	0%	50%
MH: Injuries	25%	30%
MH: Vascular Diseases	25%	20%
MH: Renal Diseases	0%	20%
Medication	50%	80%
Lab	25%	60%
Immobility/Bed Rest	25%	70%
Demographics	50%	90%
Physical Examination	50%	70%
Procedures	25%	70%

Conclusion

The majority of CDSS studies indicate a clinically strong positive effect by managing to significantly reduce pulmonary embolisms. Greater coverage of VTE-related item categories could be observed in outcome-improving CDSS compared to outcome-neutral studies. Therefore, a comprehensive set of input variables should be considered for future re-implementation. Input data element forms were standardized and semantically enriched with UMLS codes and are available as download in different formats for reuse.

Acknowledgements

The work was funded by the German Ministry for Education and Research. Grant ID: 01ZZ1602B.

References

- [1] B. Chaudhry, J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, et al, Systematic review: impact of health information technology on quality, efficiency, and costs of medical care, *Ann. Intern. Med* **144**(10) (2006), 742-752.
- [2] B. Lindblad, N. Sternby, and D. Bergqvist, Incidence of venous thromboembolism verified by necropsy over 30 years. *BMJ* **302**(6778) (1991), 709-711.
- [3] G.I. Nazarenko, E. B. Kleymenova, S.A. Payushik, V.A. Otdelenov, D.A. Sychev, and L.P. Yashina, Decision support systems in clinical practice: the case of venous thromboembolism prevention, *Int J Risk Saf Med* **27** Suppl 1(2015), S104-105.
- [4] N. Kucher, S. Koo, R. Quiroz, J.M. Cooper, M.D. Paterno, B. Soukonnikov, et al, Electronic alerts to prevent venous thromboembolism among hospitalized patients. *NEJM* **352**(10) (2005), 969-977.
- [5] K. Karnik, FDA regulation of clinical decision support software, *J Law Biosci* (2014), doi:10.1093/jlb/lsu004.
- [6] European Commission, Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory. MEDDEV 2.1/6, (2012).
- [7] A. Hamad, I.G. Cavel, J. Hinton, P. Wade, and C. Whittlesea, A pre-postintervention study to evaluate the impact of dose calculators on the accuracy of gentamicin and vancomycin initial doses, *BMJ Open* **5**(6) (2015), e006610.
- [8] M. Dugas, Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems, *Methods Inf Med* **53**(6) (2014), 516-517.
- [9] S. Kraus, M. Enders, H. Prokosch, I. Castellanos, R. Lenz, and M. Sedlmayr, Accessing complex patient data from Arden Syntax Medical Logic Modules, *Artif Intell Med* (2015), doi: 10.1016/j.artmed.2015.09.003.
- [10] E. Shortliffe, R. Davis, S. Axline, B. Buchanan, C. Green, and S. Cohen, Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system, *Comput Biomed Res* **8**(4) (1975), 303-320.
- [11] J. Varghese, M. Kleine, S. Gessner, S. Sandmann, and M. Dugas. Effect of clinical decision support systems on patient outcomes in inpatient settings: a systematic review. PROSPERO: International prospective register of systematic reviews. PROSPERO 2016:CRD42016049946 Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016049946.
- [12] D. Moher, A. Liberati, J. Tetzlaff, and D. Altman, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, *Ann. Intern. Med* **151**(4) (2009), 264-269.
- [13] B. Shea, C. Hamel, G. Wells, L. Bouter, E. Kristjansson, J. Grimshaw, et al, AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews, *J Clin Epidemiol* **62**(10) (2009), 1013-1020.
- [14] J. Varghese, Online Supplemental Material. Available on: <https://unimuenster.sciebo.de/index.php/s/QnXqjbyJP8gaNEe>. Accessed in Dec 2016.
- [15] S. Jones, R. Rudin, T. Perry, and P. Shekelle, Health information technology: an updated systematic review with a focus on meaningful use, *Ann. Intern. Med* **160**(1) (2014), 48-54.
- [16] A. Hoerbst, W. Hackl, N. de Keizer, H. Prokosch, M. Hercigonja-Szekeres, and S. de Lusignan, Exploring complexity in health: an interdisciplinary systems approach, *Stud Health Technol Inform*, (2016), IOS Press: 978-1-61499-677-4.
- [17] D. Hunt, R. Haynes, S. Hanna, and K. Smith, Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review, *JAMA* **280**(15) (1998), 1339-1346.
- [18] Agency for Healthcare Research and Quality, *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*, Agency for Healthcare Research and Quality Rockville, MD, 2011.
- [19] Operational Data Model. Specified by Clinical Data Interchange Standards Consortium.. Available from: <http://www.cdisc.org/odm>. Accessed 2016 Oct. .
- [20] M. Dugas, P. Neuhaus, A. Meidt, J. Doods, M. Storck, P. Bruland, et al, Portal of medical data models: information infrastructure for medical research and healthcare. Database, *The Journal of Biological Databases and Curation* (2016), bav121.
- [21] J. Varghese, C. Holz, P. Neuhaus, et al, Key data elements in myeloid leukemia, *Stud Health Technol Inform* **228** (2016), 282-286.
- [22] M. Daabiss, American Society of Anaesthesiologists physical status classification, *Indian J Anaesth* **55**(2) (2011), 111-115.
- [23] S.W. Lin, W.Y. Kang, D.T. Lin, J. Lee, F. L. Wu, C.L. Chen, et al, Comparison of warfarin therapy clinical outcomes following implementation of an automated mobile phone-based critical laboratory value text alert system, *BMC Med Genomics* **7** Suppl 1 (2014), S13.
- [24] Medical Data Models for CDSS input. Available from: <https://medical-data-models.org/welcome/search?title=%22Input+Data+Elements+CDSS%22>. Accessed 2016 Oct. .
- [25] J.D. Mitchell, J.F. Collen, S. Petteys, and A.B. Holley, A simple reminder system improves venous thromboembolism prophylaxis rates and reduces thrombotic events for hospitalized patients, *JTH* **10**(2) (2012), 236-243.
- [26] C. Williams, Y. Asi, A. Raffenaud, M. Bagwell, and I. Zeini, The effect of information technology on hospital performance, *Health Care Manag Sci* **19**(4) (2016), 338-346.
- [27] S. Hopewell, K. Loudon, and M. Clarke, Publication bias in clinical trials due to statistical significance or direction of trial results, *Cochrane Database Syst Rev* **21**(1) (2009).
- [28] D. Blumenthal. Launching HITECH, *NEJM* **362**(5) (2010), 382-385.
- [29] J. Varghese, S. Schulze Sunninghausen, and M. Dugas. Standardized cardiovascular quality assurance forms with multilingual support, UMLS coding and medical concept analyses, *Stud Health Technol Inform* **216** (2015), 837-841.

Address for correspondence

Julian Varghese. Email: Julian.Varghese@ukmuenster.de

CD-KES: An Ontology Based Knowledge Education System for Patients with Chronic Diseases and Its Constructing Approach

Chengkai Wu^a, Ling Gou^a, Yu Wang^a, Yong Shang^a, Yu Tian^a, Jing-song Li^a

^aEngineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou Zhejiang Province, China

Abstract

Patients' participation plays a crucial role in the management of chronic diseases. Educating patients about their diseases allows patients to self-regulate their daily health conditions more reasonably and effectively. This study focuses on an informative way to develop daily education and guidance among patients. We provide a systematic approach to establish, process, and present the knowledge map of chronic diseases. An ontology technique is used to model clinical knowledge and rules. Rule-based inference service is constructed based on the RETE reasoning algorithm. Several considerations in semantic visualization and interaction are listed as recommendations. A prototype of Chronic Disease Knowledge Education System (CD-KES) based on this approach has been built for diabetes mellitus. With the prototype, comparative evaluations for system performances in knowledge querying and browsing are taken. The results show the system can help patients to more easily understand medical knowledge and avoid potential negative consequences.

Keywords:

Chronic disease; Education of patients; Decision support systems

Introduction

Resulting from the large change in modern lifestyle, the global population with chronic diseases increases rapidly. Taking diabetes mellitus as an example, according to an estimation made by the International Diabetes Federation (IDF), 8.2% of adults aged 20-79 (387 million people in total) around the world were living with diabetes in 2014. In 2035, this number may rise beyond 592 million [1]. Diabetes, as well as many other chronic diseases, often lead to serious complications such as diabetic retinopathy and coronary diseases. Many of the complications may turn into long-term diseases if not treated in their early phases. Even so, an estimated ratio of global undiagnosed diabetes in adults is 46%, suggesting millions of people are unaware of their increased risk of diabetes-related complications [2].

Patients with chronic diseases and related complications require a comprehensive therapy, including daily treatments in medicine, exercise, and diet. To maintain a well-organized therapy, patients have to pay a lot for drugs, medical instruments, and advisors. In 2014, the estimated global average health expenditure of diabetic patients ranges from \$1,583-2,842 USD [3]. Despite the financial burden, the time cost for both patients and doctors to arrange and monitor daily treatments is also very high. With the lack of medical care providers, it is hard for doctors to analyze all pieces of daily data coming from each patient.

A reasonable means to improve treatment efficiency and decrease time redundancy in healthcare is to educate patients with the knowledge of diseases. Enough knowledge storage allows patients to self-regulate diseases more positively and comprehensively. Traditional ways to pass medical knowledge include giving oral advice or delivering information booklets. Despite having been implemented for decades, these doctor-led educating procedures have several disadvantages [4]. In many cases, patients are often told only a small amount of caring knowledge, which doctors think to be most relevant to their symptoms. The effort to educate patients may also fail due to the lack of updated knowledge which suits newly occurred complications or body conditions.

Patients' lack of an overall understanding of chronic diseases may cause serious consequences. When new symptoms show up, some patients may search online for treatment methods, which can be misleading such as drug advertisements or unproven therapies. Other patients may just ignore the newly grown complications, which leads to severe treatment delay that risks their lives. Therefore, it is important and meaningful to construct an accurate, explicit and flexible knowledge education system in which patients can conveniently learn, query and be notified about their most needed information.

The purpose of this paper is to introduce a system called Chronic Disease Knowledge Education System (CD-KES) and its systematic approach to establish, process, and present the knowledge map of chronic diseases. As a classic chronic disease, diabetes mellitus was selected to be the object of this study. An ontology model that contains diabetes-related knowledge resources was constructed. Rules based on standard clinical guidelines were settled to enrich semantic content and provide inference service. With the ontology based knowledge model and a flat designed user interface, patients would have flexible options to browse diabetes-related knowledge nodes and query treatment recommendations precisely.

Comparative evaluations of system performance on knowledge querying and browsing were also done. The query result was compared with the result gathered from search engines. The knowledge browsing ability was evaluated by comparing with an ontology visualization plugin OntoGraf. Evaluation results showed that CD-KES has a better performance on chronic disease knowledge educating.

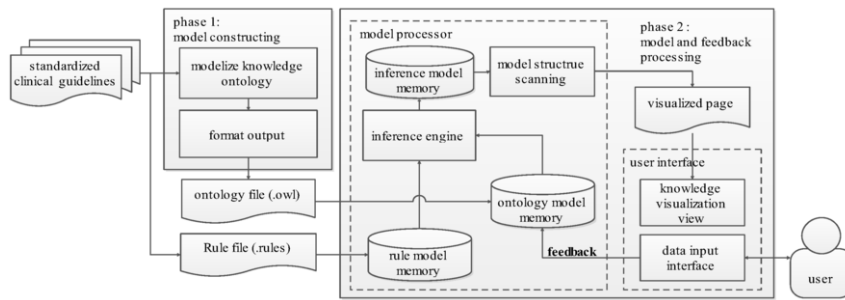


Figure 1 – Flowchart for Chronic Disease Knowledge Education System

Methods

The workflow contains procedures of disease modelization, inference rule mapping, semantic reasoning, model visualization, and personalized query service. These procedures can be separated into two phases based on the feedback loop shown in Figure 1. With this division, the procedures in phase 1 are processed beforehand and remain static during the interaction. Oppositely, procedures in phase 2 will circularly process user's input personal data and refresh query results.

Modelization and Rule Mapping with Ontology Technique

In phase 1, the knowledge model and inference rules of diabetes mellitus are mapped from standardized clinical guidelines to ontology resources. Selected guidelines include classification, diagnosis standard, treatment recommendations of diabetes from WHO [5], ADA [6] and drug labels in FDA website. The mapped knowledge resources cover the main fields of personal details, general body findings, examinations, diseases, drugs, exercise, diet, units, usages, etc.

To support the inference service and maintain a flexible resource connection, we chose an ontology-based model structure. The ontology model of diabetes is stored in OWL (Web Ontology Language) files. OWL is a standard language recommended by W3C for ontology construction and storage. It is a higher-level language of RDF (Resource Description Framework) and expanded the framework with restraints such as disjoint and equivalent. These expansions made ontology more explicit and suitable for reasoning.

An ontology model contains resources of type Class, Property, and Individual. The constraints of a superclass will be inherited by its subclasses and individuals. A definitive resource of 'diabetic_medicine' is created with expressions in Figure 2.

```
<rdf:Description rdf:about="diabetic_medicine">
  <rdf:subClassOf rdf:resource="drug"/>
  <SNOMED_ID rdf:datatype="string">R-409C4</SNOMED_ID>
  <rdf:type rdf:resource="Class"/>
</rdf:Description>
```

Figure 2 – Way to define ontology resource

The expressions respectively announced this resource

1. is a subclass of class 'drug'
2. has a SNOMED-id of R-409C4
3. is a class resource

Despite ontology resources, triples of [object, property, subject] are stored to reflect relations in the real world. To fit the knowledge model with real medical conditions, we should control resources and triples to follow sets of constraints.

Though OWL provides inner axioms, several requirements are noted to be satisfied during modelization.

1. A class of 'patient' is placed outside knowledge base to simulate patients that have differed disease details.
2. Properties connected to 'patient' class cannot have other domains.
3. Properties not connected to 'patient' class cannot add new triples during rule reasoning.

With the above requirements and limitations, this knowledge model will remain static during reasoning and query processes. All properties are separated into patient properties and knowledge properties, so it is also constrained to reduce potential semantic conflicts, as what Figure 3 shows.

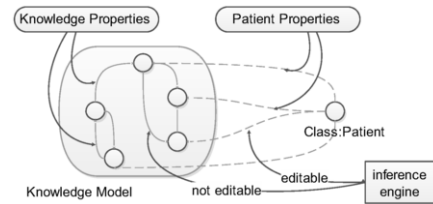


Figure 3 – Restraints in knowledge modeling

Rules express logical causality with format [name: conditions → inferred results]. When all conditions of a rule expression are satisfied, the inference engine will execute the inferred results. Rules are set to uncover indirect knowledge relations inside the ontology model. As an example, the following rules in Figure 4 are used to filter out forbidden drugs based on patients' physiologic details.

```
[Drug01: (?p DM:hasFoundDisease ?e)
  (?e DM:hasForbiddenDrugs ?d)
  -> (?p DM:hasForbiddenDrugs ?d)]

[Drug02: (?p DM:hasBodyFindings ?e)
  (?e DM:hasForbiddenDrugs ?d)
  -> (?p DM:hasForbiddenDrugs ?d)]

[Drug03: (?p DM:hasPregnancy DM:pregnant)
  (?d DM:hasPregnancyWorry true)
  -> (?p DM:hasForbiddenDrugs ?d)]
```

Figure 4 – Way to define inference rules

Model Reasoning with RETE Algorithm

The inference engine in this study uses RETE algorithm to optimize reasoning efficiency. RETE was firstly developed by Charles Forgy. The core idea of the RETE algorithm is to decompose and interconnect elements in sets of rules to reduce redundant in data storage and calculation consumption [7]. Each rule can be separated into five types of nodes to share common components between rules, namely Root Node, Type

Node, Pattern Node, Join Node, and Terminal Node [8]. The optimized network is called a ‘genetic’ rule. To demonstrate the mechanism, we transferred the rules named ‘Drug02’ and ‘Drug03’ in Figure 4 into a genetic rule shown in Figure 5.

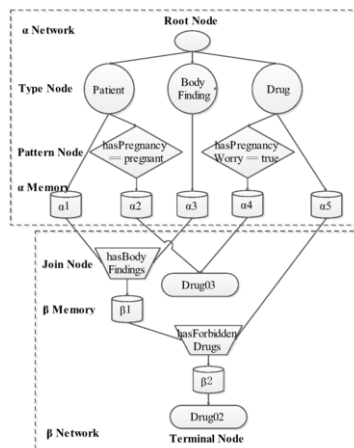


Figure 5 – Inference mechanism for RETE algorithm

A genetic rule network contains two sub-networks of α and β networks to support literal and inter-condition restrictions.

In α network, ontology resources are sent to the Root Node when reasoning process begins. Type Nodes are applied to filter out resources that are not of required types. In this case, these required types are defined as domains and ranges of each property. Filtered resources are sent to Pattern Nodes to have constant value tests. All resources that satisfy restrictions in Type Nodes and Pattern Nodes will be listed in α memories.

In β network, lists in α memories are joint by Join Nodes. Members from left and right lists are paired and Join Nodes will judge whether each pair fits the stored inter-condition restrictions. Each Join Node creates a new list of resource pairs and stores it in a β memory. The list will be connected to several following Join Nodes as left lists. After all lists in α memories being joint together, a final list will be sent to Terminal Node, which represents the exit of a rule. Terminal Node then executes the inferred result in rule expressions such as changing status or appending additional triples.

An important feature of the RETE algorithm is the support for increment operation. Once an inference procedure finishes, inference results including final and partial results will still be stored in α and β memories for the possible repeating reasoning request. When any modification on resources or triples occurs, only the modified part needs to re-reason. Typically, if no change takes place in α memories, no further process or time delay happens in β network.

Since the diabetes ontology model is knowledge-based, which is much smaller than most databases, the memory redundancy in this case to support the RETE algorithm can be ignored for modern computers. It is also important to discuss whether the algorithm fits the system's requirements. As discussed above, it is a limitation that only triples that have the specific domain as ‘patient’ class can be generated or edited. The knowledge model always remains static during query services, and so is the memory to store them. With a small amount of increment, the use of the RETE can shorten the time-consuming in query service and be a reasonable and ideal selection in this study.

Model Scanning and Visualization

Knowledge visualization is a young research area established on the foundation of data visualization and information

visualization. Different from data or information, knowledge contains more abstract descriptions with full semantic structures. Therefore, Knowledge visualization focuses more on the direct and explicit presentation of concepts, experience, attitudes, expectations, opinions, advices or predictions. It is also important in visualization to help acceptors better reconstruct and utilize knowledge.

Knowledge visualization based on semantic structure is called Semantic visualization. It is common to visualize a semantic model itself in semantic visualization processes, for the structure already contains huge semantic details. Typically, to visualize ontology models, the semantic elements of properties, triples, and rules should be emphasized.

With differed user demand, semantic visualization can be implemented in three approaches [9].

1. Hierarchical semantics visualization
2. Relational semantics visualizations
3. Entity-based semantics visualization

The three approaches focus on knowledge structure, context, and isolated word respectively. In our study, the approaches are performed as a combination to offer query services in varies scales. Though it is very flexible to build the visualization view, based on the features of patients such as the possible amateur background, we offer several recommendations that can help to maximum the knowledge display ability during semantic visualization.

1. The hierarchy of the knowledge model needs to be emphasized by visual elements such as size, color, shape, etc. Medical knowledge is more similar to a tree structure rather than graph. So that a display of hierarchy helps patients build an overall impression.
2. Every resource in ontology model needs to be presented and be focused as the central node. In this way, the context of each word or object can be fully expressed.
3. The description of each node and relation needs to be given in both structure context and natural language. This helps nodes to explain itself and patients can easier understand it.

The core consideration in visualization view design is providing limitless scales for patients to browse the knowledge model. In our design, patients can go through the model from the very root to any arbitrary knowledge node. The context is shown in forms of relations. Once patients select an individual, all related knowledge nodes will be displayed as well.

Several guiding measures are designed to help understand and manage the knowledge map, such as navigations and tooltips. User interfaces to upload personal and examination details are provided. Drop-down menus and check boxes are used to simplify user operation in query service. The figure of UI is presented in Figure 6 in the result section.

Evaluation of System Performance

The purpose of this study was to offer daily education and advice services for patients. In this section, we designed two evaluation methods to determine system performance, including querying performance and browsing performance. The evaluation methods are applied to the prototype of CD-KES with a theme of diabetes mellitus.

To simulate patients' possible daily action, we designed a scenario where a patient finds an abnormal test result from his blood glucose meter at home. To provide more specific settings for this scenario, we gave two sub-scenarios with body conditions of ‘mild renal insufficient’ and ‘pregnant’ respectively. We recruited six participants with no medical

background and randomly divided them into two groups. All of the participants are between age 20 to 24 and had experience using the internet for more than 5 years. Also, none of the participants had been renal insufficient or pregnant. Participants in each group role-played a patient with a FBG (fasting blood glucose) value of 11.0 mmol/L and one of the additional conditions above. Participants were told to freely, but independently search for online medical information for 20 minutes using search engines such as Google. Their searching results were compared with the query result from CD-KES to evaluate the efficiency and accuracy of our system.

Table 1 – Test scenario for querying performance evaluation

Scenario No	Symptom 1	Symptom 2	Replication Number
1	FBG=11.0 mmol/L	Mild renal insufficiency	3
2	FBG=11.0 mmol/L	pregnant	3

Knowledge browsing performance is also a major indicator to evaluate the knowledge display ability of the system, especially the visualization part. In this part, the visualization view of the system was compared with views generated by OntoGraf plugin. We recruited 10 participants and divided them into two groups. We randomly selected five triples from the ontology model and provided them to all participants. Participants in one group needed to find the visualization elements representing the triples in OntoGraf view, while the other group used the view of CD-KES. Time costs and click times were measured to analyze system's knowledge browsing performance. Since the evaluation focused on the browsing behavior, search bars in each view were banned.

Results

Query Evaluation Result

To have a clear evaluation standard, though diet and exercise

In both tables, single plus (+) symbol represents a drug that is recommended, single minus (-) symbol represents potentially forbidden drug, double minus (--) symbol represents clear forbidden drug. If two symbols are listed in the same cell, it suggests the advice differs with the specific drugs.

Browsing Evaluation Result

The visualization view for knowledge display and query is shown in Figure 6.

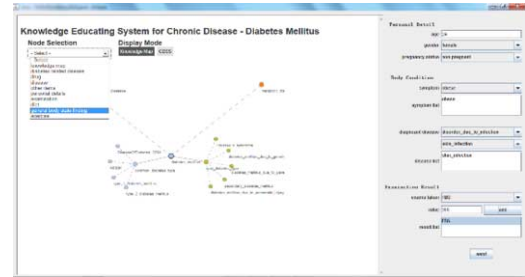


Figure 6 – Visualization view of CD-KES

Browsing performance is shown in Figure 7. This figure shows the differences in time cost for participants to find all five triples in both views.

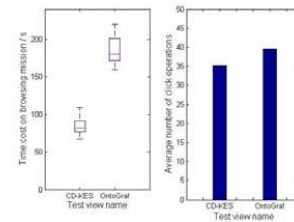


Figure 7 – Evaluation result for browsing performance

Discussion

Table 2– Test result for 1st scenario (diabetes with mild renal insufficiency)

Result From	DPP4 inhibitor	Alpha glucosidase inhibitor	human insulin analog	Biguanide	Sulfonylurea	incretin mimetic
System	-	+	+	--	--(-)	-
Patient 1			+		+	
Patient 2	+	+	+	--		+
Patient 3			+	--	+	

Table 3– Test result for 2nd scenario (diabetes with pregnancy condition)

Result From	DPP4 inhibitor	Alpha glucosidase inhibitor	human insulin analog	Biguanide	Sulfonylurea	incretin mimetic
System	-	-	+(+)			-
Patient 1	-	-	-	-	-	-
Patient 2	-	-	+	-	-	-
Patient 3	-	-	-	-	-	-

results are also offered from our system and online advice, only six types of diabetic drugs were compared in this section.

In the case of diabetes patient with mild renal insufficiency, the query result and participants' online search results are displayed and compared in Table 2. For patients with pregnancy condition, the results are displayed and compared in Table 3.

Discussion for Query Evaluation Result

In the case of diabetes patients with mild renal insufficiency, our system gave two kinds of recommended drugs and two kinds of forbidden drugs. However, participants were recommended to use additional kinds of drugs such as the DPP4

inhibitor, Sulfonylurea or incretin mimetic products when they tried to get online advice.

A possible explanation is that participants didn't notice whether one kind of drugs was useful for type-1 diabetes or type-2 diabetes. For example, DPP4 inhibitors and incretin mimetic are ineffective to type-1 diabetes. Since no personal detail was input about what kind of diabetes the patients had, our system gave out recommendations suitable for both kinds of diabetes. While a few of Sulfonylurea drugs like gliquidone, are harmless for patients with mild renal insufficiency, many of the drugs in this class are still forbidden base on drug labels. Some websites provided participants with misleading information which could have resulted in serious consequences.

In the case of diabetes, patients with the pregnant condition, two of the three participants refused to use any drugs. They were convinced that the safest way to treat diabetes during pregnancy is just to eat less fruit and sweets.

However, it has been proven that most of the human insulin analogs are harmless for pregnant women. More importantly, the possible harm caused by using drugs such as biguanide or sulfonylurea will be much less than what could be caused by untreated diabetes. Blindly staying away from all drugs may cause treatment delay and hurt both the mother and the baby.

Therefore, the results given by our system are more reasonable and comprehensive than the ones gathered from the internet. It also shows that CD-KES has the potential to educate patients with disease knowledge, help patients to better self-regulate their daily treatments, and reduce possible misleading information.

Discussion for Browsing Evaluation Result

It is hard to evaluate the performance of visualization views; and there is no golden standard for quantification. In this case, however, the time cost for participants to find triples in OntoGraf view was almost twice of that cost given by CD-KES. Though the time cost varied, the click numbers made by participants were close. The two measures suggested that participants with OntoGraf spent more time on going through knowledge nodes to find the correct ones. This result partly explained the differences in explicitness between two views that CD-KES performed better in knowledge exhibition. Since OntoGraf is still a wild-used plugin in the field of semantic visualization, it shows plenty of requirements for better visualization methods.

Conclusions

The aim of this study was to find an effective way to educate and guide patients with chronic diseases in their daily life. The article provides a knowledge education system called CD-KES and its systematic approach to establish, process, and present knowledge of chronic diseases. Evaluating the performance of this approach, a prototype system of CD-KES, with a scenario of diabetes mellitus is built. The evaluation results shows that this approach is reasonable and can largely reduce possible misleading information or treatment delays. It also has the potential to help patients self-regulating diseases and cut expenditures.

Acknowledgements

This work was supported by Chinese National High-tech R&D Program (No.2015AA020109), National Key Scientific Instrument and Equipment Development Project (No.2016YFF0103200), and the Fundamental Research Funds for the Central Universities of China.

References

- [1] L. Guariguata, D.R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J.E. Shaw, Global estimates of diabetes prevalence for 2013 and projections for 2035, *Diabetes Res ClinPr* **103** (2014), 137-149.
- [2] J. Beagley, L. Guariguata, C. Weil, and A.A. Motala, Global estimates of undiagnosed diabetes in adults, *Diabetes Res Clin Pr* **103** (2014), 150-160.
- [3] J.d.R. Fernandes, K. Ogurtsova, U. Linnenkamp, L. Guariguata, T. Seuring, P. Zhang, D. Cavan, and L.E. Makaroff, IDF Diabetes Atlas estimates of 2014 global health expenditures on diabetes, *Diabetes Res ClinPr* **117** (2016), 48-54.
- [4] R.L. Street and P. Haidet, How Well Do Doctors Know their Patients? Factors Affecting Physician Understanding of Patients' Health Beliefs, *J Gen Intern Med* **26** (2011), 21-27.
- [5] K. Alberti, P.Z. Zimmet, and W.H.O. Consultation, Definition, diagnosis and classification of diabetes mellitus and its complications part 1: Diagnosis and classification of diabetes mellitus - Provisional report of a WHO consultation, *Diabetic Medicine* **15** (1998), 539-553.
- [6] A. American Diabetes, Standards of medical care in diabetes-2015 abridged for primary care providers, *Clinical diabetes: a publication of the American Diabetes Association* **33** (2015), 97-111.
- [7] C.L. Forgy, RETE: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem, *Artif Int* **19** (1982), 17-37.
- [8] A.F. Ronszcka, R.F. Banaszewski, R.R. Linhares, C.A. Tacla, P.C. Stadysz, J.M. Simao, and Ieee, Notification-Oriented and Rete Network Inference: A Comparative Study, in: *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 2015, pp. 807-814.
- [9] K. Nazemi, D. Burkhardt, E. Ginters, and J. Kohlhammer, Semantics Visualization - Definition, Approaches and Challenges, in: *2015 International Conference Virtual and Augmented Reality in Education*, E. GonzalezMendivil, P.G.R. Flores, J.M. Gutierrez, and E. Ginters, eds., 2015, pp. 75-83.

Address for correspondence

Corresponding author: Jing-song Li.

Address: College of Biomedical Engineering and Instrument Science, Zhejiang University, 38 Zheda road, Hangzhou 310027, China.

Email: ljs@zju.edu.cn.

Tel.: +86-571-87951564

Gathering Real World Evidence with Cluster Analysis for Clinical Decision Support

Eryu Xia^a, Haifeng Liu^a, Jing Li^a, Jing Mei^a, Xuejun Li^b,
Enliang Xu^a, Xiang Li^a, Gang Hu^a, Guotong Xie^a, Meilin Xu^c

^a IBM Research – China, Beijing, China

^b Department of Endocrinology and Diabetes, the First Affiliated Hospital, Xiamen University, Xiamen, Fujian, China

^c Pfizer Investment Co. Ltd., Beijing, China

Abstract

Clinical decision support systems are information technology systems that assist clinical decision-making tasks, which have been shown to enhance clinical performance. Cluster analysis, which groups similar patients together, aims to separate patient cases into phenotypically heterogeneous groups and defining therapeutically homogeneous patient subclasses. Useful as it is, the application of cluster analysis in clinical decision support systems is less reported. Here, we describe the usage of cluster analysis in clinical decision support systems, by first dividing patient cases into similar groups and then providing diagnosis or treatment suggestions based on the group profiles. This integration provides data for clinical decisions and compiles a wide range of clinical practices to inform the performance of individual clinicians. We also include an example usage of the system under the scenario of blood lipid management in type 2 diabetes. These efforts represent a step toward promoting patient-centered care and enabling precision medicine.

Keywords:

Cluster Analysis, Expert Systems, Software

Introduction

Precision medicine has gained increasing attention as an innovative and revolutionizing approach of healthcare that takes individual traits into account in order to develop more effective ways to improve health and treat disease [1]. Unlike traditional medical treatments designed for the ‘average patient’ or ‘standard patient,’ precision medicine tailors medical decisions, practices, interventions, and products to individual patients based on their predicted response to treatment or risk of disease. While precision medicine aims at providing the best available care for each patient based on disease subclasses within a disease of common biological basis, the discovery of such subclasses, and the translation of this knowledge into clinical practice, depend critically on information technology [2]. One possible solution for subclass discovery is via cluster analysis, which fulfills the task of grouping a set of objects in a way that objects in the same group (or cluster) are more similar to each other than to those in other groups. Patient grouping via cluster analysis takes a solid step towards precision medicine by separating patients into phenotypically heterogeneous groups and defining therapeutically homogeneous patient subclasses. Patient clusters such as diagnosis-related groups and ambulatory patient groups were traditionally used for management purposes by identifying iso-resource groups (reviewed in [3]). Cluster analysis was also used for more accurate phenotyping of heart failure and related syndromes [4, 5]. A study of geriatric stroke patients using cluster analysis [6] represents an

attempt to identify meaningful patient clusters for developing specific treatment programs.

Clinical Decision Support Systems (CDSS) are health information technology systems that assist clinical decision-making tasks. Studies in CDSS emphasize improvements in system quality, including enhancing clinical performance for drug dosing, preventive care, and other aspects of medical care. The effectiveness of CDSS in improving patient outcomes is mixed [7]. While traditional CDSS was based on domain knowledge, e.g., digitized clinical guidelines, real world evidence (RWE) has gained more importance in CDSS with advances in data accumulation and technology. RWE is the acquisition of evidence from data derived from medical practices among heterogeneous sets of patients in actual practice settings [8]. Cluster analysis of clinical data represents a class of methods to generate RWE. Though cluster analysis has demonstrated its usefulness towards personalized medicine, its potential has not been fully utilized in CDSS.

In order to take advantage of cluster analysis to further enhance clinical performance and improve patient outcomes using CDSS, we have implemented a system that uses cluster analysis for CDSS with the following steps. First, cluster analysis is conducted following the clustering pipeline. Statistical methods can then be applied to evaluate the clusters and discover group profiles. After deciding on the clusters, cluster representatives will be selected and used in assigning new patient visits into existing clusters. This assignment is automatically conducted in the background and is updated upon changes to the patient record. The assigned cluster’s profile will then be available for clinicians to provide evidence and support for clinical decision-making tasks upon new patient encounters. This work represents an integration of cluster analysis in clinical decision support systems, and has the potential to promote patient-centered care and enable precision medicine.

Methods

The method is organized into three parts:

1. Description of the working pipeline of the system, which includes the initial cluster analysis, statistical analysis after achieving clusters, cluster representative selection, and cluster assignment upon new patient visits
2. Description of the use of cluster analysis for CDSS
3. Description of a working example

Cluster analysis

A typical clustering workflow, as shown in Figure 1, includes three major steps:

1. Pre-clustering processing

2. Clustering
3. Post-clustering evaluation

Pre-clustering processing consists of feature engineering work including extracting features from raw data, imputing missing data, normalizing continuous features, selecting features relevant for the working task, and reducing dimensionality, all of which serve the purpose of increasing data quality and suitability for cluster analysis. Clustering methods such as hierarchical clustering, centroid-based clustering including K-means, model-based clustering, and density-based clustering like DBSCAN, have all been used in healthcare settings to separate patients into different clusters. The exact use of methods depends on the data characteristics, the requirements of the task, and user preference. Post-clustering evaluation includes internal evaluation and external evaluation with a different validation data source, data type, and evaluation objectives. Internal evaluation of sample separation is easy to conduct but is less effective at improving information retrieval and addressing clinical implications. External evaluation with known class labels or external benchmarks is ideal for assessing clustering quality, but these benchmarks are scarce in healthcare datasets. At times, statistical analysis described in the next step is used as a compromise.

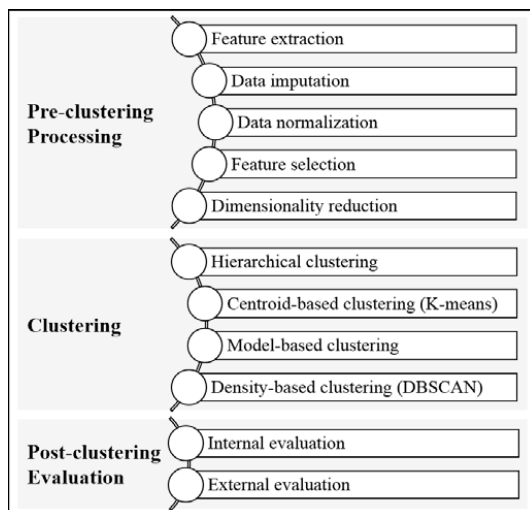


Figure 1. A typical clustering workflow

Statistical analysis

After cluster analysis, statistical analysis should be conducted to describe the profile of the clusters from three perspectives:

1. Summarize features with highest clinical relevance within each cluster; this generates the cluster profile
2. Extract distinct features of each cluster to provide justification and support for the clustering result
3. Summarize the representative action or response of each cluster with respect to the application

As an example, under a drug recommendation scenario, within each cluster, features with clinical relevance such as gender, age, disease stage or severity, and major disease history would be summarized as the cluster profile. Later, from each cluster, distinct features like 'older than 60 years old', 'diabetes' would be extracted as distinguishing features of this cluster to justify the underlying clustering rationale. Then, the drug usage would

be summarized within each cluster to provide evidence for drug recommendation.

Apart from assisting in drug recommendation, statistical analysis can also provide insights into the common diagnosis of the cluster, the typical disease onset and progression of the cluster, the general disease prognosis or clinical outcomes of the cluster, the typical response to a clinical practice and drug effectiveness within the cluster. For example, patients within a cluster may have an especially high occurrence of a disease, poor survival rate, or a good response to a certain drug.

Cluster representative selection

Cluster representatives are selected for two purposes:

1. To enable quick cluster assignment upon new patient encounter by comparing only against the selected representatives instead of exhaustively against all patient cases
2. To avoid comparisons against actual patient information, which can lead to a potential violation of privacy

Some clustering methods have nature cluster representatives, e.g., centroids in centroid-based clustering like K-means, mean values in Gaussian mixture model-based clustering. Representatives can also be defined in other clustering methods using mean values, or as the points with the highest kernel density. The determined cluster representatives can then be used by the system to provide decision support.

Cluster assignment and background update

Because patient information and status changes over time, the patient information at a specific time is called a patient case in this work. The cluster analysis was conducted at a specific time, thus representing a clustering of patient cases. At the time of use, the patients may not have been used for cluster analysis or may have information changed. Before first use, the similarity between each patient and each cluster representative is calculated and stored at the time of initialization. Cluster assignment is then conducted based on the clustering algorithm and the selected representatives. The distance between the patient case and each cluster representative is calculated. The patient case is assigned to the cluster whose representative is most similar. This assignment is automatically conducted in the background and is updated upon patient record changes. The similarity between the patient case and all cluster representatives is updated incrementally based on the recorded changes made without complete recalculation.

Cluster analysis for CDSS

After cluster assignment, the assigned cluster's profile as achieved in the statistical analysis is available for the clinicians to provide evidence and support for clinical decision-making tasks upon new patient encounters. In the current scenario, given a patient record, treatment would be automatically recommended based on digitized clinical guidelines. The cluster profile is provided as real world evidence for decision support. The clinician can view the cluster profile, including summaries of clinically relevant features like gender, age, disease stage, or treatment history. The common diagnosis, typical disease onset and progression, general disease prognosis or clinical outcomes, typical response to clinical practice, drug effectiveness and usage within the cluster can also be viewed. The clinician can then leverage the information in his/her own prescription.

Working example

An example dataset was used to demonstrate the usage of the system, which is a collection of type 2 diabetes patient records

across multiple hospitals in a city of China. The working scenario of the system is for blood lipid management in type 2 diabetes patients. In this demonstration, the system was constructed following the workflow mentioned above, and used to provide evidence for guideline-based drug recommendation.

Results

A working example is shown in this section to demonstrate the use of the workflow and system in CDSS without special elaboration on and care for the construction and evaluation of the clustering pipeline or the scientific insights. The example is concerned with blood lipid management in type 2 diabetes patients where patients with high blood lipid level and satisfying certain conditions would be recommended to use statins to lower blood lipids per the clinical guideline. The usage of statins in the cluster that the new patient belongs to would thus be shown as real-world evidence for prescription. Since elevated blood lipid level is a well-established risk factor for cardiovascular disease, decrease in cardiovascular disease onset would be used to show the effectiveness of statin treatment.

Cluster analysis

The example cohort includes patients with type 2 diabetes. A patient case is generated each time a patient visits the hospital and obtains prescriptions. The patient's clinical data from the preceding year is considered. A total of 187115 cases were defined. Features used for cluster analysis include demographics that were stratified and one-hot encoded, diagnosis history given by the International Classification of Diseases (ICD) code and counted, and treatment history as the number of times each medicine was prescribed. These altogether resulted in 1446 features. Mini Batch K-means was used with Euclidean distance to cluster the samples into 1000 clusters using sklearn [9]. The distribution of cluster size is shown in Figure 2 as a density plot (x-axis denoting the cluster size and y-axis denoting the corresponding density), where the majority of clusters have less than 100 patients and a few large clusters exist. Cluster representatives are defined as the centroids generated during Mini Batch K-Means clustering.

Statistical analysis was first conducted to summarize cluster features such as patient gender, age group, diabetes disease stage, hypertension, and hyperlipidemia in comparison with all patients. The usage of statins was also summarized at different levels: using statins or not, distribution of the statins' generic names among patients using statins, and distribution of the statins' brand name for each generic name.

Since elevated blood lipid levels are well-established risk factors for cardiovascular diseases, decrease in cardiovascular disease onset would thus be used to show the effectiveness of statin treatment. The response to statin usage within each cluster was calculated by stratifying the patients into two groups, those using statins and those not using statins, within which the number of patients later diagnosed with cardiovascular diseases were counted. Odds ratio is calculated to measure the association between the exposure (the usage of statins) and the outcome (cardiovascular disease onset) in each patient cluster. The odds ratios were plotted as dots in Figure 3, where the error bars show the 95% confidence intervals (x-axis representing the cluster index organized by increasing odds ratio and y-axis denoting the value of odds ratio or confidence interval). As shown, different patient clusters have different odds ratios compared to the odds ratio in the whole patient cohort (denoted by the red horizontal line). This provides evidence that the clusters have different responses to statin

treatment and justification for the differential use of statins in different patient clusters.

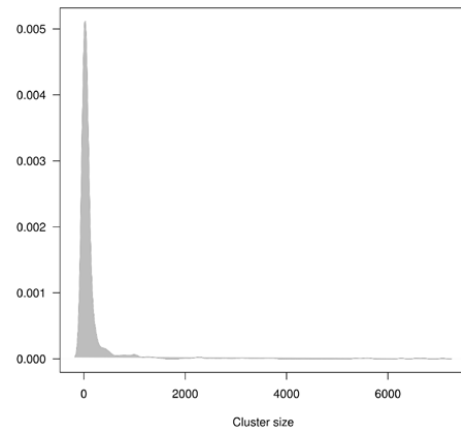


Figure 2. Distribution of cluster size

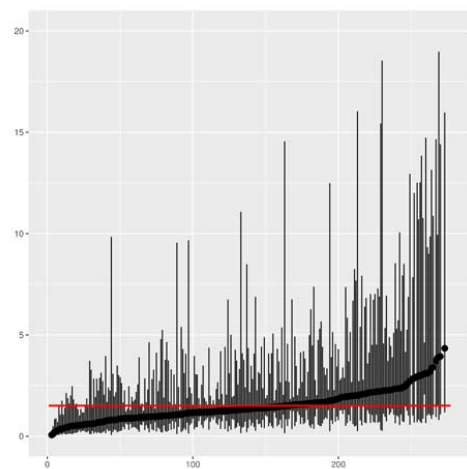


Figure 3. Cluster response to statin treatment

Cluster assignment and background update

Euclidean distance was used as the distance measure in the previous clustering steps. As a result, patient cases were assigned to the cluster whose centroid is the shortest Euclidean distance away. The centroids from the cluster analysis would be stored with the system. During initialization, the similarity between each patient and each centroid would be calculated and stored, and each patient would be assigned a cluster following the above method. Afterwards, the cluster assignment would be updated in the background each time the patient record changes. A patient record change would require an update of the patient's similarity to each centroid incrementally without completely recalculate the distances. Means of the background update of Euclidean distance are summarized in Table 1. In our working example, the features are either categorical or the number of times an event happens. The change in the squared distance can be easily calculated using Equations (1) to (3). Only features that have been changed were included in the calculation, where the contribution to squared distance change were calculated

respectively. In Table 1, v_j denotes the value of a centroid at a specific feature, v_i denotes the previous value of the patient at this feature, while v_i' denotes the new feature value. After summing up all the squared value changes, the new distance (d'_{ij}) would be calculated from the original distance (d_{ij}) using Equation (8). Similarity value here is calculated as one minus the distance.

Table 1– Squared distance change calculation

Categorical feature	Dichotomous	
	Nominal	$\Delta d_{ij}^2 = (-1)^{v_i' \neq v_j} = v_j$ (1)
	Nominal (One-hot)	$\Delta d_{ij}^2 = 2 \times (-1)^{v_i' \neq v_j}$ (2)
	Ordinal	
Continuous feature		$\Delta d_{ij}^2 = (v_i' + v_i - 2 \times v_j) \times (v_i' - v_i)$ (3)
Aggregate feature	Mean	$m_i' = \frac{m_i \times T_i + v_i'}{T_i + 1}$ (4)
	Max	$m_i' = \max(m_i, v_i')$ (5)
	Min	$m_i' = \min(m_i, v_i')$ (6)
		$\Delta d_{ij}^2 = (m_i' + m_i - 2 \times m_j) \times (m_i' - m_i)$ (7)

$$d'_{ij} = \sqrt{d_{ij}^2 + \Delta d_{ij}^2} \quad (8)$$

We also show in Table 1 the update of aggregate features. In Equation (4), m_i denotes the mean value before change, T_i denotes the number of records before change, and v_i' denotes the newly added value. The mean value after change (denoted by m_i') can thus be calculated. The squared distance change can then be calculated using Equation (7), and the new distance can be calculated with Equation (8). Similar calculations can be done with (5) for maximum value, with (6) for minimum value, where m_i indicates the maximum value and minimum value respectively, each followed by calculation with (7) and (8).

The similarity values between the patient and each centroid can thus be updated. The patient cluster assignment changes if the most similar centroid changes. Centroids themselves do not change during system usage.

Cluster analysis for CDSS

The working example is about an application of cluster analysis for blood lipid management in type 2 diabetes patients. Statins are a group of cholesterol-lowering drugs. If elevated blood lipid levels are observed upon patient encounter, statins may be recommended for use based on clinical guidelines. At this time, a clinician may turn to the CDSS for evidence of statin usage among similar patients, i.e. patients in the same cluster. Screenshots of the system are shown in Figure 4.

The usage of statins in the patient's assigned cluster would be used as evidence for prescription. In Figure 4, a page is shown with drug recommendations for blood lipid management, where statins and fibrates (medication that lowers blood triglyceride levels, always used in combination with statins, not elaborated on here) are listed in two independent tabs. In the screenshot, the total number of patients in this cluster is 2329. Of these, 1113 used statins (1031 used atorvastatin and 82 used rosuvastatin). After clicking the bar for atorvastatin or rosuvastatin, the distribution of brand names used are shown in the second and third plot. From Figure 4, we see that nearly half of similar patients used statins, where the majority of the statin users used atorvastatin, the major brand name being Atorvastatin Calcium Tablet (Lipitor). This thus provides the clinician with evidence from real-world data.

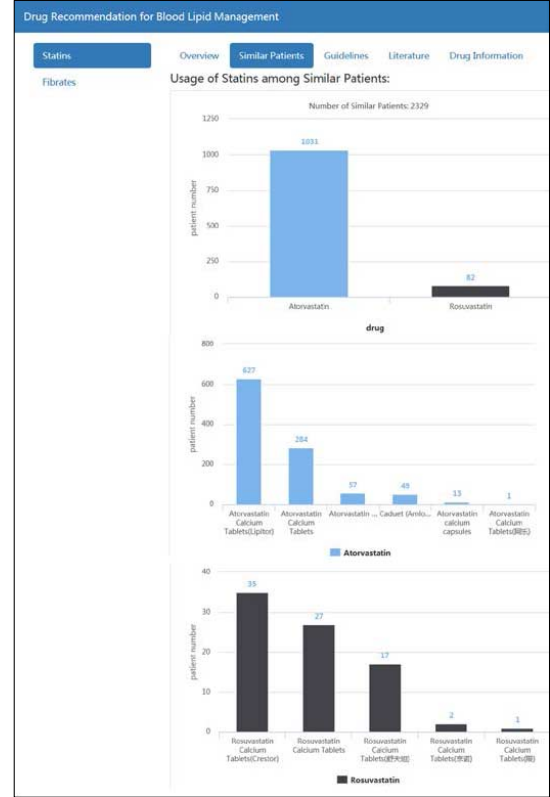


Figure 4. Screen shot of the CDSS system

Discussion

We see the intuition of using cluster analysis in CDSS from two perspectives. First, both patient and disease heterogeneity are inherent, which divides patients into distinct clusters that have different clinical features such as disease onset, disease prognosis, or drug effectiveness. Second, doctors naturally make decisions based on experiences summarized from previous similar cases. In an effort to simulate human cognitive processes and provide more extensive and deeper insights, clustering is conducted to standardize the definition of 'similar patients.' Information on the previous treatment of a large population can be summarized and integrated for the doctor's use. This system makes use of cluster analysis to provide real-world evidence for clinical decision support.

The system as a framework is useful, while the decision support aspect relies heavily on the data quality and cluster analysis pipeline adopted by the user. Many methods and approaches are available for use in all three steps of cluster analysis. However, as a class of unsupervised methods, cluster analysis cannot be easily assessed in terms of clustering quality and suitability for the application objective. This problem gets exacerbated in healthcare settings where data quality is not always satisfactory, and external benchmarks or data with known class labels are rarely available. Statistical analysis is thus included in the workflow after cluster analysis to provide some assistance for interpretation. Given this limitation of cluster analysis, we apply it to provide evidence for drug recommendations instead of directly providing recommendations in the demonstration. With improvements in data quality and clustering quality, this system has the potential to be used to directly provide drug recommendations.

Clinician trust in information technologies may be a main limiting factor in the application of the system in drug recommendation even when good data and algorithms exist for cluster analysis. Here, we advise readers not to take such a system as an independent thinking entity with unknown logic, but as a means to learn and present others' experience to facilitate individual doctor practice.

Conclusions

Here, we formulated the application of cluster analysis as providing real world evidence for clinical decision support. We described the workflow and provided a description with a demonstration of the system on an example dataset. The workflow, starting with cluster analysis and statistical analysis, generates cluster representatives, against which each new patient case would be assigned a cluster. The cluster assignment is updated in the background each time patient data gets updated. This system provides clinical decision support by presenting real-world evidence for guideline-based treatment. It has the potential to promote patient-centered care and enable precision medicine.

Acknowledgements

We thank our colleagues in the Cognitive Healthcare team in IBM Research – China for the help and suggestions.

References

- [1] F. S. Collins and H. Varmus, "A new initiative on precision medicine.," *N. Engl. J. Med.*, vol. 372, no. 9, pp. 793–5, 2015.
- [2] P. N. Robinson, "Deep phenotyping for precision medicine," *Hum. Mutat.*, vol. 33, no. 5, pp. 777–780, May 2012.
- [3] H. SANDERSON and L. MOUNTNEY, "The development of patient groupings for more effective management of health care," *Eur. J. Public Health*, vol. 7, no. 2, 1997.
- [4] T. Ahmad, M. J. Pencina, P. J. Schulte, E. O'Brien, D. J. Whellan, I. L. Piña, D. W. Kitzman, K. L. Lee, C. M. O'Connor, and G. M. Felker, "Clinical implications of chronic heart failure phenotypes defined by cluster analysis," *J. Am. Coll. Cardiol.*, vol. 64, no. 17, pp. 1765–1774, 2014.
- [5] S. J. Shah, D. H. Katz, S. Selvaraj, M. A. Burke, C. W. Yancy, M. Gheorghiade, R. O. Bonow, C. C. Huang, and R. C. Deo, "Phenomapping for novel classification of heart failure with preserved ejection fraction," *Circulation*, vol. 131, no. 3, pp. 269–279, 2015.
- [6] B. I. Buijck, S. U. Zuidema, M. Spruit-van Eijk, H. Bor, D. L. Gerritsen, and R. T. C. M. Koopmans, "Is patient-grouping on basis of condition on admission indicative for discharge destination in geriatric stroke patients after rehabilitation in skilled nursing facilities? The results of a cluster analysis.," *BMC Health Serv. Res.*, vol. 12, no. 1, p. 443, 2012.
- [7] A. X. Garg, N. K. J. Adhikari, H. McDonald, M. P. Rosas-Arellano, P. J. Devereaux, J. Beyene, J. Sam, and R. B. Haynes, "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review," *JAMA*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [8] "Real World Evidence: A New Era for Health Care Innovation," 2015.
- [9] XiaF. Pedregosa and G. Varoquaux, *Scikit-learn: Machine learning in Python*, vol. 12. 2011.

Address for correspondence

Eryu Xia, PhD, IBM Research – China, 19A, Zhongguancun Software Park, Haidian District, Beijing, China. Email: xerbj@cn.ibm.com

Context-Sensitive Clinical Alert Packages Written in Arden Syntax

Julia Zeckl^{a,b}, Katharina Adlassnig^b, Renate Fossler^c, Alexander Blacky^d,
Jeroen S. de Bruin^{b,e}, Walter Koller^f, Andrea Rappelsberger^e, Klaus-Peter Adlassnig^{b,e}

^a University of Applied Sciences Technikum Wien, Höchstädtplatz 6, A-1200 Vienna, Austria

^b Medexter Healthcare GmbH, Borschkegasse 7/5, A-1090 Vienna, Austria

^c Vienna Regional Health Insurance Fund, Karl-Aschenbrenner-Gasse 3, A-1210 Vienna, Austria

^d VAMED-KMB Hospital Management and Operation GmbH, Spitalgasse 23, A-1090 Vienna, Austria

^e Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

^f Department of Hospital Epidemiology and Infection Control, Medical University of Vienna and Vienna General Hospital, Währinger Gürtel 18–20, A-1090 Vienna, Austria

Abstract

An increasing body of raw patient data is generated on each day of a patient's stay at a hospital. It is of paramount importance that critical patient information be extracted from these large data volumes and presented to the patient's clinical caregivers as early as possible. Contemporary clinical alert systems attempt to provide this service with moderate success. The efficacy of the systems is limited by the fact that they are too general to fit specific patient populations or healthcare institutions. In this study we present an extendable alerting framework implemented in Arden Syntax, which can be configured to the needs and preferences of healthcare institutions and individual patient caregivers. We illustrate the potential of this alerting framework via an alert package that analyzes hematological laboratory results with data from intensive care units at the Vienna General Hospital, Austria. The results show the effectiveness of this alert package and its ability to generate key alerts while avoiding over-alerting.

Keywords:

Decision Support Systems, Clinical; Laboratory Critical Values; Infection Control

Introduction

Given the increasing body of raw clinical data being provided in electronic medical record (EMR) systems, a quality control mechanism is needed to ensure that potentially critical patient information is extracted from these large data volumes and offered to the appropriate patient caregivers in a timely manner. Clinical alerts are a part of such information; they make the patients' caregivers aware of immediate pathological, unusual, or occasionally life-threatening circumstances concerning the patient.

Electronic storage of patient data creates opportunities for computerized monitoring of patients and subsequently computerized generation of clinical alerts. Over the years, a substantial number of monitoring systems have been developed for a variety of healthcare settings, especially in the fields of infection control and adverse event detection [1–5]. The benefits of such systems are manifold: they offer complete and timely information on patients who require the immediate attention of caregivers, errors can be prevented, patient safety is enhanced, and institutional quality standards for patient care are upheld or improved.

A drawback of these systems is that their embedded detection rules or classification methods are usually based on guidelines

for “standard” cases. As such, the generated monitoring results and alerts may be valid only for patients without additional underlying conditions. For example, de Bruin et al. describe an automated surveillance system for monitoring healthcare-associated infections [6]. In their study, a false-positive infection episode was detected due to a large number of leukocytes. However, the underlying cause of the leukocyte abnormality was not an infection but the patient's leukemia.

Modification of existing systems is not a viable option. There is usually little room for system or module customization; the systems are provided “as-is”. This limits their use in healthcare environments other than the one(s) they were developed for. Individual institutions or even departments within an institution may employ different guidelines or different interpretations of the same guideline. A configurable and extendable framework of alert systems was developed by the use of established standards for the creation and integration of clinical decision support systems (CDSSs). The framework can be used directly at the point of care, and is tailored to the needs and preferences of healthcare institutions, departments, or even individual caregivers.

In this paper we present a preliminary version of such a framework for clinical alerts. Using Arden Syntax [7], an HL7 International [8] standard for computerized representation and processing of medical knowledge, we created an alert framework that supports configuration according to departmental or institutional requirements, such as the configuration of critical value limits, optimization of alert frequency to prevent over-alerting, and adaptable alert timing. Using patient data obtained from the Vienna General Hospital (VGH), Austria, we describe the potentialities of this alerting framework through an implemented alert package for the analysis of hematological laboratory test results. By way of an example, time- and context-sensitive clinical alerts were constructed for two infection parameters: *C-reactive protein (CRP) concentration* and *leukocyte count*. The alert package provides alerts for “standard” situations, as well as a variety of nonstandard contexts, such as underlying leukemia or preexisting infection episodes. Furthermore, alerts can be configured to the preferred frequency, thus avoiding excessive alerting.

Table 1 – Definitions of clinical alerts in an alert package for the analysis of hematological laboratory test results, including the respective alert description and context

Alert variable	ID	Message	Context	Rule
CRP concentration				
	C1	Slightly increased CRP	Previous value in the normal range or no previous value	$20 \text{ mg/l} \leq \text{CRP concentration} < 50 \text{ mg/l}$
	C2	Moderately increased CRP	Previous value in the normal range or no previous value	$50 \text{ mg/l} \leq \text{CRP concentration} < 100 \text{ mg/l}$
	C3	Significantly increased CRP	Previous value in the normal range or no previous value	$\text{CRP concentration} \geq 100 \text{ mg/l}$
	C4	Further increase of CRP	No leukemia, 4 th day of infection and beyond, value available between 12 and 36 hours prior to the current value.	$\text{CRP concentration} \geq 20 \text{ mg/l}$ and $\text{CRP concentration} - \text{yesterday's CRP concentration} \geq 0.20$ yesterday's CRP concentration
Leukocyte count				
	L1	Leukocyte value indicates leukopenia	-	$0.5 \text{ G/l} < \text{leukocyte count} \leq 3 \text{ G/l}$
	L2	Leukocyte value indicates leukocytosis. Admission diagnosis: leukemia	Leukemia	$\text{Leukocyte count} \geq 12 \text{ G/l}$
	L3	Significant increase of leukocytes	-	$\text{Leukocyte count} - \text{previous leukocyte count} \geq 0.4$ previous leukocyte count
	L4	Significant decrease of leukocytes	-	$\text{Leukocyte count} - \text{previous leukocyte count} \leq -0.4$ previous leukocyte count
	L5	Increased leukocytes compared to previous finding	No leukemia, no previous infection, previous value in the normal range.	$\text{Leukocyte count} \geq 12 \text{ G/l}$
	L6	Persistent leukocytosis	No leukemia, 4 th day of infection and beyond, value available between 12 and 36 hours prior to the current value	$\text{Leukocyte count} \geq 12 \text{ G/l}$ and $\text{leukocyte count} - \text{yesterday's leukocyte count} \geq -0.10$ yesterday's leukocyte count

	04/17/2013 13:02	04/16/2013 12:52	04/15/2013 11:29	04/14/2013 13:37	04/13/2013 13:46
HEMATOLOGICAL PROFILE					
Leukocytes	19.4 G/L	20.1 G/L	17.5 G/L	17.1 G/L	–
Hemoglobin	–	–	–	–	–
...					
Inflammation markers					
C-reactive protein	169 mg/l	105 mg/l	98 mg/l	80 mg/l	–
...					
BIOCHEMICAL PROFILE					
Electrolytes					
Potassium	–	–	–	–	–
...					
Kidney function					
Blood urea nitrogen	–	–	–	–	–
Serum creatinine	–	–	–	–	–
Urea	–	–	–	–	–
...					
Enzymes					
Troponin I	–	–	–	–	–

Clinical Alerts	
GENERATED	MESSAGES
04/17/2013 13:02	Further increase of CRP.
	Further increase of CRP (169 mg/l, 04/17/2013 13:02) compared to previous finding (105 mg/l, 04/16/2013 12:52). CRP 169 mg/l (04/17/2013 13:02) CRP 105 mg/l (04/16/2013 12:52) CRP 98 mg/l (04/15/2013 11:29) CRP 80 mg/l (04/14/2013 13:27)
04/17/2013 13:02	Persistent leukocytosis.
	Persistent leukocytosis (19.4 G/L, 04/17/2013 13:02) compared to previous finding (20.1 G/L, 04/16/2013 12:52). Leukocytes 19.4 G/L (04/17/2013 13:02) Leukocytes 20.1 G/L (04/16/2013 12:52) Leukocytes 17.5 G/L (04/15/2013 11:29) Leukocytes 17.1 G/L (04/14/2013 13:27)
04/16/2013 12:52	Further increase of leukocytes.
04/14/2013 13:27	Leukocyte value indicates leukocytosis.
04/14/2013 13:27	Moderately increased CRP.

Figure 1 – The HTML5 web application frontend. On the left side, the patient's laboratory values for C-reactive protein (CRP) and leukocytes are displayed for the present day and three preceding days; values in red are associated with a clinical alert. On the right side, clinical alerts are displayed together with a short description. On clicking the alert, a more detailed alert message is shown along with an overview of the laboratory values related to the alert over the past few days. Alerts shown in bold type are those for the present day.

Methods

Outcome and variables

The primary outcome measures were the frequency and nature of the alerts generated by the alert package for each of the aforementioned clinical variables.

We used patient demographics (age, sex, length of stay) to describe the patient population. Laboratory data for the clinical variables *CRP concentration* (mg/l), and *leukocyte count* (G/L) were used for the generation of alerts. We report on the number of generated alerts for each alert rule.

Study design, setting, and participants

A retrospective single-center cohort study was performed on prospectively collected and validated data at VGH, a 1,933-bed tertiary-care and teaching hospital, and was approved by the ethics committee of the Medical University of Vienna. All adult patients (i.e., age ≥ 18 years) admitted to a VGH intensive care unit (ICU) for at least 24 hours between 1 January and 31 December 2013 were eligible for the study. Patients for whom laboratory values for both clinical variables (CRP concentration and leukocyte count) were not available were excluded from the study.

Data management and sample size

Demographic patient data as well as laboratory test results were retrieved from the Philips IntelliSpace Critical Care and Anesthesia information system, which is in operation at almost every ICU of the VGH.

The interrogation of data sources using the previously mentioned selection criteria yielded a total of 266 patient stays comprising 2,830 patient days.

Knowledge base and data processing

The alert package discussed in this paper is part of a framework consisting of automated, context-sensitive and customizable alert packages targeting a variety of issues in clinical routine. They were created by repeated discussions with experienced

clinical experts and by eliciting their feedback on a variety of use cases in which alerts were generated for different clinical situations. Based on these discussions and feedback, alert rules were implemented by clinical knowledge engineers that, based on laboratory test results for the aforementioned clinical parameters, generate one or multiple (different) alerts. We discuss ten alerts that were defined in the knowledge base: four for CRP levels and six for leukocyte count. These alerts, together with their context and corresponding rules, are shown in Table 1.

The alerts were implemented in a knowledge base with Arden Syntax. The latter is a programming language for the collection, description, and exchange of medical knowledge in a computer-executable format. For this project we used Arden Syntax version 2.10 [7]. An Arden Syntax knowledge base comprises a set of programming units known as medical logic modules (MLMs) [9]. In all 17 MLMs were constructed; 14 MLMs for data import and preprocessing, and three MLMs for alert generation for a variety of contexts.

Implementation, management, testing, and execution of the MLMs were done using the ARDENSUITE clinical decision support technology platform, which comprises an integrated development and test environment (IDE), as well as the ARDENSUITE server to execute MLMs [10]. The ARDENSUITE server can be accessed through web-service protocols, i.e., representational state transfer (REST) [11] or simple object access protocol (SOAP) [12].

Data presentation

For data presentation, an HTML5 web application was created using the Ionic framework [13] (Figure 1). Upon selecting a patient in the web application, the necessary MLMs are accessed using JavaScript JQuery [14] commands through REST calls. The ARDENSUITE server extracts parameter values from the POST REST call and executes the called MLM(s) with these parameters. The alerts generated by these MLMs are then transmitted back to the web application through a JavaScript Object Notation (JSON) [15] object and displayed in the web application.

Results

Of the 266 patients included in the study, 115 were female (43%). The patients' mean age was 61 years, with an interquartile range (IQR) of 21 years. The mean duration of the hospital stay was 11 days, with an IQR of 11 days.

In all 5,492 data entries were registered for the two clinical variables: 2,697 entries (49%) for CRP levels and 2,795 for leukocyte counts. Values for both variables were available for 2,662 patient days (94%), only CRP levels were available for 35 patient days (1%), and only leukocyte counts for 135 patient days (5%).

During the study period a total of 2,382 alerts were generated for CRP levels and leukocyte count, amounting to an average of 0.89 alerts per patient day. The maximum number of alerts displayed on a single patient day was four; the most numerous alerts displayed for a single patient during his/her stay was 90, over a period of 81 days.

CRP levels were alerted on 459 occasions (19.3%), and leukocyte counts on 1,923 occasions. Table 2 shows the number of times each alert listed in Table 1 was generated.

Table 2 – Frequency of clinical alerts generated during the study period for each alert defined in Table 1.

Alert ID	#Generated
C1	133
C2	84
C3	79
C4	163
L1	95
L2	1,032
L3	214
L4	88
L5	161
L6	333
Total	2,382

Discussion

We present a configurable framework for clinical alerts implemented in Arden Syntax. Using the framework, healthcare institutions, departments, and patient caregivers can adapt alerts to make them more relevant and useful, and avoid over-alerting. Alert fatigue is a very serious problem in clinical routine. Various studies on alerting suggest that by far the large majority of alerts are ignored or overridden [16, 17]. Most patient caregivers (about 90%) do not respond to all alarms, but rather match their response rates to the expected probability of true alarms [18]. In view of these facts, it would be very important to use an alerting framework that can be configured to the preferences of a healthcare institution or caregiver. Last but not least, this would maximize the caregiver's perception of the usefulness of alerts.

Through cooperation with clinical experts, it became evident that there is no real objective metric to establish when an alert is justified or not. Rather, there are some guidelines to when an alert is useful, namely, when it can be acted upon by some kind of intervention. Furthermore, there are some expectations as to the number of generated alerts; most physicians do not expect

more than one alert per person per patient day. With a maximum of four alerts per day and an average of 0.89 alerts per patient day, the alert package discussed here proved to be quite conservative in its alerting frequency.

Of all the alerts, L2 was generated significantly more often (in all 1,032 times; see Table 2) than the others. This may be partly explained by the fact that the alert rule is very straightforward and therefore more likely to be generated. However, the L2 alert was only meant for patients with leukemia, which makes it less likely than many others. MLMs and data consultation disclosed that the leukemia diagnosis of a patient was not included in the data transfers from VGH. Therefore, the check as to whether the patient had leukemia was not conducted by the MLM; a leukocyte count of ≥ 12 G/l was used as the sole criterion. This resulted in the alert being generated for all patients with a leukocyte count ≥ 12 G/l and not adhering to the specific context of L2, which also explains the large number.

The limitations of this study are worthy of mention. First, as this was a pilot feasibility study, we conducted a retrospective cohort study on prospectively collected data. As such, prospective behavior in this framework still needs to be studied. Furthermore, as the pilot study was conducted on a single-center basis, the generalizability of these results needs to be reviewed as well. A third limitation is the lack of other relevant patient data. Although there are alerts especially for patients with leukemia, the (admission) diagnosis for leukemia by itself is regrettably not part of the data volume to distinguish between infection-triggered leukocytosis and leukemia. Finally, we need to assess the ease and practicality of adapting MLMs to caregivers' preferences while adhering to institutional guidelines. Although the implemented alert package has been tested and adapted by clinical partners with diverse clinical backgrounds, a more extensive study with data outside the ICU setting and with multiple alert packages will be carried out for a better analysis of the framework.

Conclusion

We presented a configurable framework for automated alerting based on electronic patient data. We showed the feasibility and potentialities of the system with respect to its configuration for different contexts and the optimization of alert frequency.

References

- [1] J.S. de Bruin, W. Seeling, and C. Schuh, Data Use and Effectiveness in Electronic Surveillance of Healthcare Associated Infections in the 21st Century: A Systematic Review, *J Am Med Inform Assoc* **21** (2014), 942-951.
- [2] R. Freeman, L.S.P. Moore, L. García Álvarez, A. Charlett, and A. Holmes, Advances in Electronic Surveillance for Healthcare-Associated Infections in the 21st Century: A Systematic Review, *J Hosp Infect* **84** (2013), 106-119.
- [3] S.M. Handler, R.L. Altman, S. Perera, J.T. Hanlon, S.A. Studenski, J.E. Bost, M.I. Saul, and D.B. Fridsma, A Systematic Review of the Performance Characteristics of Clinical Event Monitor Signals in the Hospital Setting, *J Am Med Inform Assoc* **14** (2007), 451-458.
- [4] J. Leal and K.B. Laupland, Validity of Electronic Surveillance Systems: A Systematic Review, *J Hosp Infect* **69** (2008), 220-229.
- [5] R.A. Raschke, B. Gollihare, T.A. Wunderlich, J.R. Guidry, A.I. Leibowitz, J.C. Peirce, L. Lemelson, M.A. Heisler, and C. Susong, A Computer Alert System to Prevent Injury From Adverse Drug Events: Development and Evaluation in a Community Teaching Hospital, *JAMA* **280** (1998), 1317-1320.

- [6] J.S. de Bruin, K.-P. Adlassnig, A. Blacky, H. Mandl, K. Fehre, and W. Koller, Effectiveness of an Automated Surveillance System for Intensive Care Unit-Acquired Infections, *J Am Med Inform Assoc* **20** (2013), 369-372.
- [7] Health Level Seven, *Arden Syntax v2.10 (Health Level Seven Arden Syntax for Medical Logic Systems, Version 2.10)*, 2015. Available at: http://www.hl7.org/implementation/standards/product_brief.cfm?product_id=372, last access: 8 November 2016.
- [8] Health Level Seven (HL7) International, <http://www.hl7.org/>, last access: 24 November 2016.
- [9] G. Hripesak, Writing Arden Syntax Medical Logic Modules, *Comput Biol Med* **24** (1994), 331-363.
- [10] Medexter Healthcare, *ArdenSuite – Medical Knowledge Representation and Rule-Based Inference Software with Arden Syntax*, 2015. Available at: <http://www.medexter.com/arden-syntax/arden-syntax>, last access: 8 November 2016.
- [11] R.T. Fielding, *Architectural Styles and the Design of Network-based Software Architectures*, Dissertation, University of California, Irvine, 2000.
- [12] R.T. Fielding, *Architectural Styles and the Design of Network-based Software Architectures*, Dissertation, University of California, Irvine, 2000.
- [13] Ionic, <https://ionicframework.com/>, last access: 24 November 2016.
- [14] JQuery, <https://jquery.com/>, last access: 24 November 2016.
- [15] JavaScript Object Notation (JSON), <http://www.json.org>, last access: 24 November 2016.
- [16] B.J. Drew, P. Harris, J.K. Zègre-Hemsey, T. Mam-mone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, and X. Hu, Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients, *PloS ONE* **9** (2014), e110274.
- [17] T. Isaac, J.S. Weissman, R.B. Davis, M. Massagli, A. Cyrulik, D.Z. Sands, and S.N. Weingart, Overrides of Medication Alerts in Ambulatory Care, *Arch Intern Med* **169** (2009), 305-311.
- [18] J.P. Bliss, R.D. Gilson, and J.E. Deaton, Human Probability Matching Behaviour in Response to Alarms of Varying Reliability, *Ergonomics* **38** (1995), 2300-2312.

Address for correspondence

Univ.-Prof. DI Dr. Klaus-Peter Adlassnig, FACMI
 Section for Artificial Intelligence and Decision Support
 Center for Medical Statistics, Informatics, and Intelligent Systems
 Medical University of Vienna
 Spitalgasse 23, A-1090 Vienna, Austria
 E-mail: klaus-peter.adlassnig@meduniwien.ac.at and
 Medexter Healthcare GmbH
 Borschkegasse 7/5, A-1090 Vienna, Austria E-mail:
kpa@medexter.com

Validation of the Chinese Version of the Functional Oral Intake Scale (FOIS) Score in the Assessment of Acute Stroke Patients with Dysphagia

Hongzhen Zhou^a, Yafang Zhu^a, Xiaomei Zhang^b

^aDepartment of Nursing, Nanfang Hospital, Southern Medical University, Guangdong, China.

^bDepartment of Neurology, Nanfang Hospital, Southern Medical University, Guangdong, China.

Abstract

This study aimed to validate the Chinese version of the Functional Oral Intake Scale (FOIS) score in acute stroke patients with dysphagia. A sample of 128 consecutive patients with acute stroke, admitted to Department of Neurology from April to October in 2016, completed the FOIS. The interrater reliability, criterion validity, discriminant validity, cross validation, and the sensitivity of FOIS scale were evaluated. Results showed that rater agreements were excellent for FOIS (Kw=0.881, $p<0.001$). A highly negative correlation between FOIS and WST (water swallow test) was detected ($r=-0.937$, $p<0.001$). There was significant difference for FOIS level of patients with different evaluation outcomes ($\chi^2=126.551$, $p<0.001$). The FOIS evaluation results were significantly correlated with two physiological measures of swallowing. The Chinese version of the FOIS score is a reliable scale for evaluating the level of oral feeding function in patients with acute stroke.

Keywords:

Weights and measures; Deglutition disorders; Stroke

Introduction

Approximately 22-65% of the patients with stroke have dysphagia [1], and it has been estimated that 40-50% of these patients encountered aspiration problems [2], dysphagia, and related complications, which will lead to prolonged hospital stay, and are associated with increased mortality, comorbidity, and health care costs [3-4]. According to the American Heart Association and American Stroke Association guidelines for early stroke management, the primary step is to screen aspiration before the administration of food, liquid or medication in stroke patients [5]. There are many bedside screening tools for aspiration in stroke patients, and the WST (water swallow test) may be the most convenient screening tool among stroke patients in nursing practice [6]. A recent systematic review has suggested that the WST have sensitivities between 64-79% and specificities between 61-81% [6]. However, the assessment of aspiration risk does not constitute the only objective in the evaluation of dysphagic stroke patients, in whom potential feeding problems should also be addressed [7]. Eating without knowledge of the dysphagia can lead not only to pneumonia, but also to life-threatening conditions, such as dehydration, malnutrition, and suffocation [8, 9]. Therefore, evaluation of oral feeding function is especially important. These concerns strongly require a reliable scale that is easy to be used, to provide further swallowing function details in feeding and nutrient intake.

Concerning this problem, the Functional Oral Intake Scale (FOIS) for dysphagia in stroke patients, a novel oral feeding

function rating system, was developed by the Florida Health Science Center in 2005 [10]. This is a 7-point ordinal scale that describes the typical functional oral intake of patients with stroke and dysphagia [11]. In the FOIS, all levels can calculate, such as what the patient consumes by mouth on a daily basis. Levels 1 through 3 are related to varying degrees of non-oral feeding, while levels 4 through 7 are varying degrees of oral feeding without non-oral supplementation, and it considers both diet modifications and patient compensations [10]. A score below 6, of a maximum of 7, indicates restrictions of oral intake of food and liquid [11]. A recent study shows that there is an association between the level of oral intake and the degree of oropharyngeal dysphagia in elderly post-stroke patients during chronic phase [12]. Another finding suggests a negative and moderate correlation between T-EAT-10 and FOIS [13]. However, this scale has been used in Japan [14] and Iran [15], but no Chinese version is available to date. The purpose of this investigation was to translate the English version of FOIS to Chinese, and to evaluate the interrater reliability, criterion validity, discriminant validity, cross-validation, and the sensitivity of the Chinese version of the FOIS in Chinese Han stroke patients with dysphagia.

Methods

This study was approved by the ethical committee of the Nanfang Hospital, Southern Medical University, Guangdong, China (No. NFEC-2016-145). Written and verbal informed consent was obtained from all patients or their caregivers, and all data were collected prospectively. A total of 128 patients with acute stroke, admitted to Department of Neurology (Nanfang Hospital, Southern Medical University) from April to October 2016 were included in this study. The inclusion criteria included willingness to participate in the study, being over the age of 18, normal cognitive function, admitted to Department of Neurology within three days of stroke onset, and having a clinical diagnosis of stroke confirmed by an attending stroke neurologist according to the World Health Organization's definition of stroke [16]. Exclusion criteria were: 1) History of other diseases that affect swallowing function, such as head and neck cancer, esophageal cancer, brain injury, myasthenia gravis or Guillain Barre Syndrome; and 2) patients with nasal feeding on admission. Demographic data, vital signs, diagnosis, day of evaluation, NIHSS, MBI, and the extent of dysphagia according to the WST were recorded.

Translation of the Chinese version of the FOIS

The items of the FOIS were first translated to Chinese by two bilingual neurology specialist, who had more than ten years of clinical experience (forward translation). A meeting, in which four dysphagia experts participated, was held to confirm a

single scale with a consensus (synthesis). Necessary adjustments in the translation were made after the consensus. Then, a native English speaker outside of the medical profession and a dysphagia expert with experience of studying abroad translated the instrument into English (backward translation). Finally, comparison by two bilingual experts derived English text and converted into a scale. The backward translation was finally sent to the original creator of the FOIS to double verified. Every item of FOIS was identical to the original version (Table 1).

Table 1 - The Chinese version of the Functional Oral Intake Scale (FOIS)

FOIS ITEMS FOIS 条目	0 = Completely consistent 0 =完全 一致 4 = Completely inconsistent 4 =完全不 一致				
Level 1: Nothing by mouth. 1级: 完全不经口进食	0	1	2	3	4
Level 2: Tube dependent with minimal attempts of food or liquid. 2级: 管饲依赖, 极少尝试进食普 通食物和液体食物	0	1	2	3	4
Level 3: Tube dependent with consistent oral intake of food or liquid. 3级: 管饲依赖, 经口进食同一质 地的普通食物和液体食物	0	1	2	3	4
Level 4: Total oral diet of a single consistency 4级: 完全经口进食单一粘稠度的 食物	0	1	2	3	4
Level 5: Total oral diet with multiple consistencies, but requiring special preparation or compensations. 5级: 完全经口进食多种粘稠度的 食物, 但需特殊制备或补给	0	1	2	3	4
Level 6: Total oral diet with multiple consistencies without special preparation, but with specific food limitations. 6级: 完全经口进食多种粘稠度食 物而无需特殊制备, 但有特殊食物 限制	0	1	2	3	4
Level 7: Total oral diet with no restrictions. 7级: 完全经口进食, 无任何限制	0	1	2	3	4

Interrater Reliability

Because of the characteristics of self-recovery of stroke dysphagia, this study did not assess the test-retest validity of the tool, but test interrater reliability to reflect the stability of FOIS. A one-page handout with written instructions describing the FOIS was provided to raters who were given opportunities to assess patients before study started. Through direct patient observation, or patient or caregiver statement, two stroke nurses, who were not involved with translation and have more than 5 years of clinical experience in neurology department, used FOIS to evaluate the amount, type and method of oral feeding or liquid to the newly admitted patients respectively. The WST was carried out according to the conventional method by 2 other stroke nurses at the time of admission. In addition, in order to minimize the possible

changes in patient's level of swallowing function, all the assessments were completed within 48 hours when no substantial change of the swallowing ability was expected to be taken place [13].

Criterion Validity

The criterion validity of the FOIS was determined by assessing the correlation between FOIS and WST, NIHSS or MBI. WST, a useful screening tool for aspiration, was used to evaluate swallowing function and reflects the severity of dysphagia. The WST has 5 levels in which level 1 means normal swallowing function, while level 5 represents severe dysphagia. In addition, studies showed that there was a significant correlation between the swallowing function and the severity of stroke [17, 18]. The National Institutes of Health Stroke Scale (NIHSS) is a validate instrument that evaluate the severity of stroke [19]. The Modified Barthel Index (MBI) was used to evaluate the patient's performance in activities of daily living (ADLs). This scale is often regarded as a functional interpretation of disability or dependency in the ADLs. The potential associations between the FOIS ratings and the measures of stroke severity, ADLs, and swallowing ability were investigated by Chi-square, Cramer's V (dichotomized data) or ϕ (multiple category data). Only when this measure was dichotomized, the obtained scores could be included in the analysis. Although no cut point on NIHSS scale is universally accepted, this value was chosen because a score of more than 8 was used in the National Institute of Neurological Disorders and Stroke recombinant tissue plasminogen activator study to define a severe post-stroke neurologic deficit [19]. Dichotomized cut-off scores were 15 for the MBI. Finally, FOIS ratings were compared with the established criteria. The above measurements were collected when the patient was admitted to a stroke unit.

Discrimination validity

Clinical comprehensive evaluation, as evaluation standard, was applied to this study [20]. On the basis of the patient's swallowing disorder and the severity of symptoms, the patients were divided into three groups: normal oral feeding group, oral feeding disorders without tube-feeding group, and tube-feeding group. The FOIS evaluation results of the three groups of patients were judged. The specific content of clinical comprehensive evaluation include the evaluation grade of WST, eating pattern and food form (exclude tooth or oral disorders), cough, whether nasal feeding, fiber optic bronchoscopy and so on. According to the actual situation of the patients and the analysis of the results of the patients' caregivers, clinical doctors and nurses, it is judged to be the existence of swallowing disorders when the patient's eating patterns and food patterns changed. In addition, cough during eating and drinking was also judged to be dysphagia.

Cross validation

Cross-validation was evaluated via comparing FOIS scores with the incidence of dysphagia and aspiration, and with the severity of dysphagia and aspiration according to video fluoroscopy swallowing study (VFSS), a golden standard for the diagnosis of dysphagia, impaired swallowing function, and aspiration [21]. All the above comparisons were completed within 72 hours of admission to the stroke unit.

Sensitivity to Change

In order to explore the sensitivity of the FOIS scale and to evaluate the changes of oral feeding function, the FOIS scale was used to evaluate the swallowing function in patients at 3 time points, at admission to the stroke unit, at 1 month post-onset, and at 3 months post-onset. Subsequently, the rating

distribution bar chart of FOIS was plotted to evaluate the changes of functional oral intake over time.

Statistical Analysis

Statistical analysis was carried out using Windows-based SPSS 20.0. Arithmetical means and standard deviations (mean \pm SD) for quantitative variables were calculated, and categorical variables were presented as frequencies. All statistical tests were conducted at a 5 % significance level.

For both the FOIS score and the WST, pairwise weighted K values were calculated. Interrater reliability was evaluated with the Cohen K statistic. A K statistic of 0.4 or less is considered poor, values between 0.4 and 0.6 are considered fair to moderate, those between 0.6 and 0.8 suggest good inter-observer agreement, and values higher than 0.8 suggest excellent agreement [22]. This approach is conservative for our reliability comparisons, in that the agreement among raters will be inflated by these automatically perfect agreements [23]. Chi-square and Cramer's V correlation analyses were calculated to assess criterion validity and cross-validation. Discriminant validity was tested by non-parametric rank sum test. And the sensitivity of the FOIS scale for clinical assessment was investigated by plotting rating distribution bar chart.

Results

Patient Characteristics

From April 2016 to October, 128 patients were enrolled. Detailed patients' characteristics are summarized in Table 2. The average age of patients was 59.16 years (median = 59 years, range = 22–80 years), and 70% were men. The diagnoses of the patients selected for the study were ischemic stroke (103 patients, 81%), hemorrhagic stroke (12 patients, 9%), unknown (13 patients, 10%). Twenty-nine patients (22%) were indwelled gastric tube.

Table 2 - Clinical Features of 128 Acute Stroke Patients

Clinical Features	Initial FOIS Ratings*				
	1	4	5	6	7
Patients	29	2	2	20	75
Mean age \pm SD (y)	59.76 \pm 13.64	74.00 \pm 7.07	69.50 \pm 7.78	60.40 \pm 8.54	57.85 \pm 10.90
Sex (%)					
Male	65.5	50	100	90	66.7
Female	34.5	50	0	10	33.3
Smoking(%)					
Yes	17.2	0	0	50	30.7
No	75.9	50	100	50	62.7
Quit smoking	6.9	50	0	0	6.7
Pathology					
Cerebral infarction	25	2	2	20	55
Cerebral hemorrhage	3	0	0	0	9
Unknown	1	0	0	0	11
Inhospital Day	12.65	12	17.5	10.95	9.08
Mean MBI score	14.66	20	37.5	60	70.80
Mean NIHSS score	10.62	6.5	8	5.15	2.59

Abbreviations: SD, standard deviation; rTPA, Recombinant Tissue Plasminogen Activator; MBI, Modified Barthel Index; NIHSS, National Institutes of Health Stroke Scale.

*No patients received FOIS ratings of 2 or 3 on the initial scoring at admission.

Interrater Reliability of the FOIS Score

The overall reliability was excellent for both the FOIS score (Kw= 0.881, Spearman, $r=0.972$) and the WST (Kw= 0.844, Spearman, $r=0.965$) (Table 3). The rater agreement was good to excellent for stroke nurses. For the FOIS score, 29 of 128 (23%) observations had the lowest level (1). The distribution of the FOIS was comparable with the distribution with the WST. One level of the WST was recorded on 81 occasions. However, 75 were scored at the highest FOIS level (7). In the remaining 9 instances, food with multiple viscosity could be taken without the help of special preparation or compensation, but some specified food could not.

Table 3 - The Kappa values and Spearman's correlation of the FOIS and The WST for the correlation between the two assessments

Rater Pair	No. of patients	FOIS			WST		
		K	Sr	P	K	Sr	P
N/N	128	.881	.972	.000*	.844	.965	.000

FOIS, Functional Oral Intake Scale; WST, Water Swallow Test; N, nurse. * $P<0.001$

Criterion Validity

The WST is a clinical screening instrument to assess the status of swallowing function in patients. As shown in the table, there is a strongest negative correlation between the FOIS and the WST ($r = -0.937$, $p < 0.001$). The NIHSS and MBI were also significantly associated with the FOIS ratings on admission to a stroke unit (Table 4).

Table 4 - Chi-Square, Spearman's correlation and Cramer's V Correlations between the FOIS Scale and the NIHSS, MBI, and WST Scale within 48 Hours of Admission to Stroke Unit

Test	X ²	P	Sr	Cramer's V Correlations/ ϕ
NIHSS	57.84	.000*	-.480	.84
MBI	61.71	.000*	-.553	.81
WST	6.18	.000*	-.937	1.73

* $P<0.001$

Discrimination validity

Clinical comprehensive evaluation was performed in all patients who met the inclusion criteria. According to the evaluation results of swallowing function, the patients were divided into three groups: normal oral feeding group, oral feeding disorders without tube-feeding group, and tube-feeding group. Non-parametric rank sum test was completed to evaluate the FOIS level among the three groups of patients. The evaluation results of the 3 groups were statistically significant difference ($\chi^2=126.551$, $P<0.001$) (Table 5), indicating the FOIS scale can be used to determine whether or not the patients had oral feeding disorder as well as the severity of the symptoms.

Table 5 - The FOIS level among the three groups of patients

Group	Patients	FOIS level	X ²	P
Normal oral feeding group	75	Level I		

Oral feeding disorders without tube-feeding group	24	Level4-6	126.55	.00
Tube-feeding group	29	Level7	1	0*

* $P < 0.001$

Cross-validation

The cross-validation analysis data showed that the FOIS was significantly associated with presence of both dysphagia and aspiration derived from VFSS. Dysphagia severity was significantly correlated with FOIS ratings ($\chi^2=65.32$; $P < 0.001$), but the aspiration severity was not (Table 6).

Table 6 - Chi-Square, Spearman's correlation and Cramer's V Correlations between the FOIS Scale and Dysphagia or Aspiration Presence or Severity

Item	χ^2	P	Cramer's V Correlations/ ϕ
Dysphagia (+)	55.96	.000*	.93
Aspiration (+)	65.48	.000*	.93
Dysphagia Severity	65.32	.000*	.52
Aspiration Severity	19.02	NS	.31

* $P < 0.001$. NS, not significant.

Sensitivity to Change

Data of oral feeding function of stroke patients at admission to a stroke unit, at 1 month and 3 months post-stroke were shown by FOIS level distribution map (Figure 1). The oral feeding function of the patients showed a gradual improvement within three months after stroke, which is consistent with the results of a previous study [24]. Therefore, the FOIS scale can sensitively reflect substantial changes of oral intake of food and liquid for stroke patients.

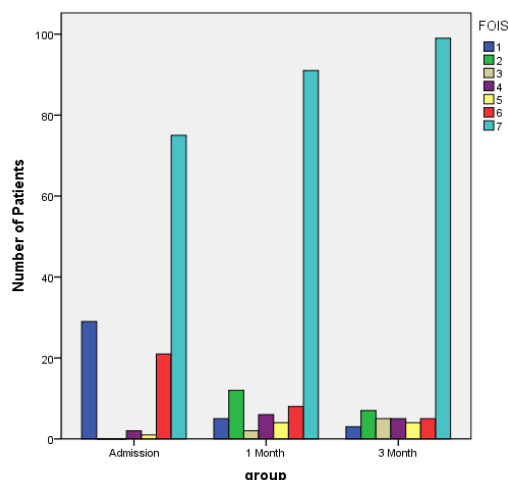


Figure 1. Changes of swallowing function in patients on admission to a stroke unit and at 1 month and 3 months post-stroke.

Discussion

As an oral feeding function evaluation tool, the FOIS scale is easily mastered and used in a diverse range of patients. It is a

clinically useful instrument to record the severity of symptoms, and to monitor the progression of the disease and the effect of treatment. It has been shown to be strongly associated with swallowing dysfunction [13, 14, 25]. What is more, it has important clinical significance to accurately record the symptoms of dysphagia and the changes of oral feeding function in patients during the whole treatment period. Therefore, it is necessary to use the FOIS scale in clinical practice.

The interrater reliability of FOIS was high, which is consistent with the results of the original authors [10], suggesting the FOIS had a high reliability and stability. In terms of validity, the results demonstrated that there was a strong negative correlation between the FOIS and WST. However, we found that participants, even in WST level 1, had a certain degree of difficulty in feeding through the mouth. For example, Level 6 refers to the patients who can eat most foods by mouth, but avoid foods that are difficult, such as meat, salad, or dry foods. The food avoidance is specifically due to swallowing difficulty. Associations between the FOIS and the NIHSS or MBI were moderate at admission to the stroke unit. The moderate correlation may be explained by the uneven distribution of patients according to the FOIS. The number of patients at the extreme ends of the FOIS was too concentrated, and there were fewer participants in middle ranges of the FOIS. Once the patients took food by detaining nasogastric tube, they would be informed that eat any food by mouth or injection of food through tubes are not allowed. This is why no patients in FOIS level 2 or 3 were included.

In addition, this study evaluated the distinguish performance of the FOIS scale by the clinical comprehensive evaluation index. The clinical comprehensive evaluation index was revised according to the literature and on opinions of neurology experts. The FOIS evaluation results of three groups patients were compared, and the difference was statistically significant ($P < 0.001$). Verified by VFSS, the FOIS shows an excellent cross-validation. Therefore, the discrimination performance of FOIS scale is excellent, and the FOIS scale can be used to determine whether or not the patients had oral feeding disorder, as well as the severity of the symptoms.

From the FOIS level distribution map, it seems sensitive to the changes in the oral feeding function of the patients. It is very important to accurately record the oral feeding function of the stroke patients throughout the whole time of their illness. We use the WST scale to evaluate the swallowing function of patients again, and combined use the FOIS scale to record the oral feeding status of the patients. In this way, doctors can have a better grasp of the patient's overall rehabilitation.

Although there was a strong correlation between FOIS and WST scale, overall, there are some subtle differences that do exist. In the study, we found that the FOIS provides more details of swallowing function than WST, and therefore, it is superior to WST due to the availability of nutritional status, and has higher ability to recognize different stages of swallowing function. Thus, we suggest that the combination of WST and FOIS scale should be employed to evaluate the swallowing functions of patients with acute stroke.

Conclusion

The Chinese version of the FOIS scale can be used to reliably assess oral feeding function in adult Chinese patients with acute stroke and is worthy of recommendation and application in clinical practice.

Acknowledgements

The study was supported by two projects of science and technology in Guangdong Province (2013B060500047) and (2014A020212542).

This study was supported by Nang Fang Hospital of Southern Medical University. Many thanks to Dr. Michael for offering us suggestions for modifications during the translation of the FOIS scale. In addition, the authors wish to acknowledge the nursing staff in the Department of Neurology for their enthusiastic participation and valuable support. Finally, the authors wish to acknowledge and thank all the patients and their families for their kind cooperation.

References

- [1] E.C. Jauch, J.L. Saver, H.J. Adams, A. Bruno, J.J. Connors, B.M. Demaerschalk, and et al, Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association *Stroke* **44**(3) (2013), 870-947.
- [2] P.E. Marik and D. Kaplan, Aspiration pneumonia and dysphagia in the elderly, *Chest* **124**(1) (2003), 328-336.
- [3] D.G. Smithard, P.A. O'Neill, R.E. England RE, and et al, The natural history of dysphagia following a stroke, *Dysphagia* **12**(1997), 188-93.
- [4] E. Brogan, C. Langdon, K. Brookes, C. Budgeon, and D. Blacker, Dysphagia and factors associated with respiratory infections in the first week post stroke, *Neuroepidemiology* **43**(2014),140-4.
- [5] S.K. Daniels, J.A. Anderson, and P.C. Willson, Valid items for screening dysphagia risk in patients with stroke: a systematic review, *Stroke* **43**(2012), 892-7.
- [6] P.C. Chen, CH. Chuang, C.P. Leong, S.E. Guo, and Y.I. Hsin, Systematic review and meta-analysis of the diagnostic accuracy of the water swallow test for screening aspiration in stroke patients, *J adv nurs* **72**(2016), 2575-86.
- [7] C.H. Marques, A.L.de Rosso, and C. Andre, Bedside assessment of swallowing in stroke: water tests are not enough, *Top stroke rehabil* **15**(2008), 378-83.
- [8] G.H. Ickenstein, J. Stein, D. Ambrosi, R. Goldstein, M. Horn, and U. Bogdahn, Predictors of survival after severe dysphagic stroke. *J neurol* **252**(2005), 1510-6.
- [9] D.G. Smithard, P.A. O'Neill, C. Parks, and J. Morris, Complications and outcome after acute stroke. Does dysphagia matter? *Stroke* **27**(1996), 1200-4.
- [10] M.A. Crary, G.D. Mann, and M.E. Groher, Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients, *Arch Phys Med Rehabil* **86**(2005), 1516-20.
- [11] M.A. Crary, G.D. Carnaby-Mann, L. Miller, N. Antonios, and S. Silliman, Dysphagia and nutritional status at the time of hospital admission for ischemic stroke *J Stroke Cerebrovasc Dis* **15**(2006), 164-71.
- [12] C.T. Mituuti, V.C. Bianco, C.G. Bentim, E.C.de Andrade, J.H. Rubo, G. Berretin-Felix, Influence of oral health condition on swallowing and oral intake level for patients affected by chronic stroke, *Clin interv aging* **10**(2015), 29-35.
- [13] N. Demir, A.S. Serel, O. Inal, and A.A. Karaduman, Reliability and Validity of the Turkish Eating Assessment Tool (T-EAT-10), *Dysphagia* **31**(2016), 644-9.
- [14] H. Shoji, A. Nakane, Y. Omosu, and et al, The prognosis of dysphagia patients over 100 years old, *Arch Gerontol Geriatr* **59**(2014), 480-4.
- [15] J. Bakhtiyari, P. Sarraf, N. Nakhostin-Ansari, A. Tafakhori, J. Logemann, S. Faghihzadeh, and M.H. Harirchian, Effects of early intervention of swallowing therapy on recovery from dysphagia following stroke, *Iran J Neurol* **15**(2015), 119-24.
- [16] Stroke -1989, Recommendations on stroke prevention, diagnosis, and therapy, Report of the WHO Task Force on Stroke and other Cerebrovascular Disorders, *Stroke* **20**(1989) 1407-31.
- [17] G. Mann and G.J. Hankey, Initial clinical and demographic predictors of swallowing impairment following acute stroke, *Dysphagia* **16**(2001), 208-15.
- [18] J.G. Heckmann, C. Stossel, C.J. Lang, B. Neundorfer, B. Tomandl, and T. Hummel, Taste disorders in acute stroke: a prospective observational study on taste disorders in 102 stroke patients, *Stroke* **36**(2005), 1690-4.
- [19] Tissue plasminogen activator for acute ischemic stroke, The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group, *N Engl J Med* **333**(1995), 1581-7.
- [20] L. Perry, Screening swallowing function of patients with acute stroke. Part one: Identification, implementation and initial evaluation of a screening tool for use by nurses, *J clin nurs* **10**(2001), 463-73.
- [21] I.J. Cook and P.J. Kahrilas, AGA technical review on management of oropharyngeal dysphagia, *Gastroenterology* **116**(1999), 455-78.
- [22] J.R. Landis and G.G. Koch, The measurement of observer agreement for categorical data, *Biometrics* **33**(1977), 159-74.
- [23] E.F. Wijdicks, W.R. Bamlet, B.V. Maramattom, E.M. Manno, and R.L. McClelland, Validation of a new coma scale: The FOUR score, *Ann neurol* **58**(2005), 585-93.
- [24] D.G. Smithard, P.A. O'Neill, R.E. England, C.L. Park, R. Wyatt, D.F. Martin, and J. Morris, The natural history of dysphagia following a stroke, *Dysphagia* **12**(1997), 188-93.
- [25] K.L. Huang, T.Y. Liu, Y.C. Huang, C.P. Leong, W.C. Lin, and Y.P. Pong, Functional outcome in acute stroke patients with oropharyngeal Dysphagia after swallowing therapy *J Stroke Cerebrovasc Dis* **23**(2014), 2547-53.

Address for correspondence

Hongzhen Zhou, email: Zhohuz2005@126.com; Xiaomei Zhang, Email: zhangxmsam@126.com.

The Impact of Censoring Drug Switching in Medication Adherence Measures of Chronic Single Ingredient Oral Drugs

Vivienne J. Zhu, MD, MS,^a J. Marc Overhage, MD, PhD,^b
Qianli Ma, MS,^c Patrick B. Ryan, PhD^d

^aMedical University of South Carolina, Charleston, SC, USA, ^bCerner, Malvern, PA, USA,

^cHumana, Louisville, KY, USA, ^dJanssen Research & Development, Titusville, NJ, USA

Abstract

We explored how drug switching impacts adherence measures for common chronic oral medications. Switching between ingredients with the same indication was detected within a 30-day grace period. The proportion of days covered (PDC) and adherent status (cutoff 0.8) for each ingredient was calculated and compared between different censoring approaches: censoring drug switching (PDCswitch), censoring the end of dispensing (PDCend), and fixed 365-day period (PDC365). Overall, 854,380 (15.9%) patients in the Optum ClinFormatics (Optum) and 150,785 (22.0%) patients in the MarketScan Multi-state Medicaid (MDCD) had at least one switch within one year. Compared with PDC365 in Optum, PDCswitch means were higher: 0.85 vs. 0.41 for antihypertensive, 0.82 vs. 0.46 for antihyperglycemics, and 0.84 vs. 0.33 for antihyperlipidemia. Further, the percentages of adherent patients were higher: 95.8% vs. 17.9% for antihypertensive, 85.5% vs. 18.9% for antihyperglycemics, and 72.1% vs. 5.3% for antihyperlipidemia. Significant and modest changes were observed between PDCswitch and PDCend.

Keywords:

Antihypertensive agents; Hypertension; Medication Adherence

Introduction

Medication adherence can be objectively assessed using prescription dispensing records from electronic databases, such as medical record systems, insurance claims, or pharmacy dispensing databases. Various medication adherence measures have been established to estimate the proportion of the number of days supply during a specified time period or during refill intervals following a drug or a set of drugs [1-3]. Several studies also identified medication discontinuation when patients failed to refill their prescription within certain number of days after exhausting a previously dispensed supply [1-2]. However, these existing measures usually do not specifically address the issue of drug switching.

Drug switching or switching between drugs, ingredients, and classes is not uncommon in patients with chronic conditions because of the side effects, perceived efficacy, cost, or patient preference [6-8]. Some studies proposed to combine the number of days of supply for all drugs or to the average adherences of all drugs when patients have been dispensed multiple drugs in a certain observation period [3-5]. However, adherence may vary between drugs that were prescribed even when they treat same disease. While there is a large literature

evaluating medication adherence of specific medical products, there is little standardization among the approaches taken to estimate adherence with an array of different assumptions being applied to administrative claims data. For example, 12 recent publications assessed the adherence of metformin or atorvastatin using either medication possession ratio (MPR) or proportion of days supplied (PDC) as the primary metric. Within these papers, researchers varied in their assumptions around the time window of interest, with some using six months and others following 12 months from initiation of therapy. Perhaps most striking was the disparity across the community in handling censoring in the adherence metric; the predominant approach was to follow patients for the full fixed time window from the index day [6-9], but four papers modified this approach to terminate the time window to when the index drug was stopped [10-13], and two papers considered switching to alternative treatments for the same indication when censoring a patient's time [14-15]. Other papers simply excluded patients with switches in therapy from consideration [16-17]. The impact of these varied assumptions has not been fully evaluated.

In this study, we develop an approach that incorporates drug switching into medication adherence measures. We then explore the difference between this approach and existing common adherence measures using dispensing records of chronic oral medications.

Methods

Data Sources and Settings

We extracted pharmacy dispensing records from Optum ClinFormatics (Optum) and MarketScan Multi-state Medicaid (MDCD), both formatted in the Observational Medical Outcome Partnership Common Data Model (OMOP CDM) [18]. The Optum and MDCD databases capture person-specific information centered on administrative information, such as age, gender, pharmacy claims data, physician and facility claims data, and lab test results data. The Optum database represents over 34 million distinct individuals from September 1, 2005 to March 31, 2012 with an average of 27 months of longitudinal observation. The MDCD database represents data on 9.6 million distinct patients from January 1, 2006 to December 31, 2009, with an average of 11 months of longitudinal observation. Source pharmacy dispensing records of Optum and MDCD were translated to the OMOP CDM drug_exposure tables, which capture inferred utilization of dispensed drugs, such as drug concept, the number of days of supply, quantity, the start date and end date of the current

dispensing. The local drug codes in the source data were mapped to the RxNorm drug concepts. All records in the drug_exposure tables were available for this study. Analyses were conducted independently for the Optum and the MDCD database.

Eligibility Criteria

The study cohort was selected from patients who were dispensed any common, single ingredient, chronic oral medications used to treat hyperglycemia, hyperlipidemia, and hypertension. These medications include sulfonylurea, biguanides, statin, beta-blocker (BB), angiotensin-converting-enzyme (ACE) inhibitor, angiotensin II-receptor antagonist (ARB), calcium-channel blocker, and Thiazolidinedione (TZD). Drug attributes (ingredient, therapeutic class, and primary indication) were identified from the OMOP vocabulary. Patients who took combination drugs were excluded. Eligible patients' prescription-filling behaviors were followed up for 365 days from the initial dispensing event of each ingredient.

Measurements

Definition of Drug Switching

Drug switching was identified when a different medication with the same indication (drug B) was initiated within 30 days of an existing medication (drug A) being terminated. At the same time, the existing medication had to be terminated within 30 days after the replacement therapy started. Switching was distinguished from the drug augmentation, which is a situation that the existing medication was not stopped within 30 days of the new drug initiation. In this study, we only probed ingredient switching within a 365-day follow-up period from the existing medication index event. Multiple switches might be identified between ingredients but only the first one was counted for this study. In the OMOP CDM, multiple dispensing records with same ingredient can be used to construct a drug era with a specified length of gap. In this study, we allowed a 0-day gap for aggregating dispensing events into a drug era, and we detected drug switching during gaps less than 30 days between drug eras. We classified the switching index into four categories according to the time when switching occurred relative to the last drug era of drug A; A) during the last era of drug A; B) at the end of the last era of drug A; C) after the last era of drug A; and D) before the last era of drug A.

Comparison of Interest

Patient medication adherence for each ingredient was measured by the proportion of days covered (PDC) following up 365 days. PDC was defined as the total number of medication-covered days divided by the number of days of the observation time period. PDC can be calculated even if a subject has only one fill and has been used increasingly to measure patient medication adherence for quality assurance [19]. Within a 365-day follow-up time period, three observation windows were applied to each ingredient: a fixed 365 days with no censoring included (PDC365), censoring based on the last dispensing date (PDCend), or censoring based on the date when a drug switches to another ingredient treating the same disease (PDCswitch). These three censoring strategies reflect the most common situations when patients change medications. Using a conventional cut-off (0.8), PDC was further classified into two levels: adherent (≥ 0.8) and non-adherent (< 0.8). Patient PDC and adherence rates were compared between different censoring approaches: censoring

switching versus censoring last dispensing date, and censoring switching versus a fixed 365 days.

Statistical Analysis

For each category of therapeutic medication, a descriptive summary of medication dispensing events, patient characteristics, and proportion of drug switching were reported. Medication adherence and adherence classification for each patient with drug switching were calculated independently for PDC365, PDCswitch, or PDCend. Paired t-tests were conducted to assess differences of patient PDC for all subjects and subgroups stratified by indication. Chi-square tests were performed to detect the difference of adherence classification between comparison groups. In addition, across all three therapeutic areas, the proportion of changes in PDC (decreased, increased, or unchanged) within each switching index category were reported. A p-value less than 0.05 were considered significant. All analyses were implemented using SAS 9.2 (SAS Institute, Cary, North Carolina).

Results

Patient Measure Summary

A total of 1,261 distinct oral single ingredient drugs for 60 ingredients (41 for hypertension, 11 for hyperglycemia, and 8 for hyperlipidemia) were identified from the OMOP vocabulary. There were 5,365,644 patients in Optum and 684,102 patients in MDCD who were dispensed at least one studied ingredient. Following up 365 days for each ingredient for each patient, 854,380 (15.9%) patients in Optum and 150,785 (22.0%) patients in MDCD had at least one switching event between ingredients. Patient demographic and switching frequencies in each therapeutic area were reported in Table 1.

Compared with PDCend, the average PDCswitch was slightly higher in antihypertensive, antihyperglycemics, and slightly lower in cholesterol lowering medications in both the Optum and MDCD databases. Consequently, adherence status ($PDC \geq 0.8$) were 5-15% higher in antihypertensive and antihyperglycemics, and about 10% lower in cholesterol lowering drugs. In Optum, antihyperglycemics had the highest percentage of patients changing adherence status (35.1%) followed by antihypertensive (30.5%), and cholesterol lowering drugs (19.2%). Similar percentages of patients who changed adherence status were observed in MDCD: 35.9% for antihyperglycemics, 37.1% for cholesterol lowering, and 26.5% for antihypertensive drugs (Table 1).

Compared with PDC365, the average PDCswitch was higher in all three drug therapeutic classes: 0.85 vs. 0.41 for antihypertensive, 0.82 vs. 0.46 for antihyperglycemics, and 0.84 vs. 0.33 for cholesterol lowering drugs in the Optum. Further, the percentages of adherent patients were much Higher: 95.8% vs. 17.9% for antihypertensive, 85.5% vs. 18.9% for antihyperglycemics, and 72.1% vs. 5.3% for cholesterol lowering drugs. Sixty-seven to 80% of patients changed adherence status; among those patients, more than 98% of patients changed from non-adherent to adherent. Similar increases were in observed in the MDCD (Table 1).

PDC Differences

Paired t-tests were conducted to compare the PDCs measured by applying different censoring approaches. In Optum, there was a significant difference between PDCswitch ($M=0.85, SD=0.17$) and PDC365 ($M=0.41, SD=0.28$); $t(1209534)=1503.2$, $p<0.0001$. Similarly, in the MDCD, there

Table 1 - Patient characteristics and medication taking behavior summary

Optum (N=5,365,644)													MDCD (N= 684,102)												
AH				AG				AL				AH				AG				AL					
total patient (n)		3,846,89				1,143,236				2,804,909				547,057				205,448				262,099			
switching																									
switch count		832,031				255,134				122,417				166,263				52,286				17,490			
Switch person count (n, %)		611,122 15.9%				189,257 16.6%				113,736 4.0%				112,004 20.5%				38,249 18.6%				15,587 5.9%			
age, year (M, SD)		55.1 11.9				53.5 10.5				54.7 9.8				51.1 13.4				50.4 11.9				52.5 11.24			
male (n, %)		331,482 54.2%				108,409 57.3%				64,720 56.9%				41,604 37.1%				12,292 32.1%				5,732 36.8%			
adherence (PDC)																									
censoring switching (M, SD)		0.85 0.17				0.82 0.17				0.84 0.17				0.83 0.19				0.82 0.19				0.81 0.21			
censoring EOD (M, SD)		0.83 0.20				0.77 0.21				0.88 0.17				0.80 0.21				0.77 0.22				0.86 0.21			
365 days (M, SD)		0.41 0.28				0.46 0.27				0.33 0.24				0.38 0.27				0.42 0.26				0.29 0.24			
adherence classification																									
censoring switching >=0.8 (n, %)		585,175 95.8%				161,844 85.5%				81,989 72.1%				108,109 96.5%				34,910 83.9%				10,932 70.1%			
censoring EOD >=0.8 (n, %)		552,200 90.4%				132,678 70.1%				94,628 83.2%				99,044 88.4%				29,510 69.4%				12,807 82.2%			
365 day >=0.8 (n, %)		109,414 17.9%				359,14 18.9%				6,052 5.3%				16,194 14.56%				8,566 14.8%				780 5.0%			
adherence status changed (censoring switch vs censoring end)																									
patient count (n, %)		186,355 30.5%				66,418 35.1%				21,891 19.2%				41,567 37.1%				13,876 35.9%				41,23 26.5%			
adherent to non-adherent (n, %)		76,690 41.2%				18,626 28.0%				17,265 78.9%				16,251 39.1%				4,238 29.8%				2,999 72.7%			
non-adherent to adherent (n, %)		109,665 58.8%				47,792 72.0%				4,626 21.1%				25,316 60.9%				9,638 70.2%				1,124 27.3%			
adherence status changed (censoring switch vs fixed 365 days)																									
patient count (n, %)		487,311 79.7%				129,650 68.5%				76,057 66.7%				93,321 83.3%				27,044 70.4%				10,184 65.3%			
adherent to non-adherent (n, %)		5,775 1.2%				186,0 1.4%				60 0.1%				703 0.8%				350 0.9%				16 0.2%			
non-adherent to adherent (n, %)		481,536 98.8%				127,790 98.6%				75,997 99.9%				92,618 99.2%				26,694 99.1%				10,168 99.8%			

Note. AH (antihypertensive); AG (antihyperglycemics); AL (antihyperlipidemia); EOD (end of dispensing event)

was a significant difference between PDCswitch ($M=0.83, SD=0.18$) and PDCend ($M=0.80, SD=0.22$); $t(236038)=60.8$, $p<0.0001$, and there was a significant difference between PDCswitch ($M=0.83, SD=0.18$) and PDC365 ($M=0.38, SD=0.27$); $t(236038)=-664.3$, $p<0.0001$. When stratified by indications, the PDCswitch values were higher in antihypertensive and antihyperglycemics, and the PDCswitch was lower in cholesterol lowering drug when comparing with PDCend (Figure 1A). Compared with PDC365, the PDCswitch values were higher (0.3-0.5) in all three subgroups (Figure 1B). We also observed approximately 10% PDCswitch values increased 0.9 compared with PDC365 in each subgroup.

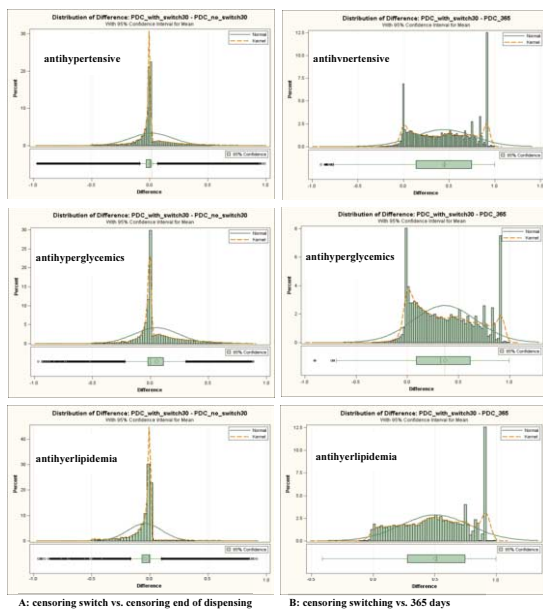


Figure 1- Distribution of PDC changes stratified by indications

Switch Index Affects Medication Adherence

In Optum, the majority of switching occurred during the last era of drug A (43.7%), followed by before the last era of drug A (30.7%), after the last era of drug A (24.3%), and at the end of the last era of drug A (1.3%). Compared with PDCend, PDCswitch might decrease, increase, or remain unchanged with different proportions (more or less) when switching occurred during or before the last era of drug A. However, PDCswitch decreased when switching occurred after the last era of drug A, and remained unchanged when switching occurred at the end of last era of drug A. Compared with PDC365, PDCswitch values were largely increased in all four switching index categories (Figure 2). Similar results were observed in the MDCD (data not shown).

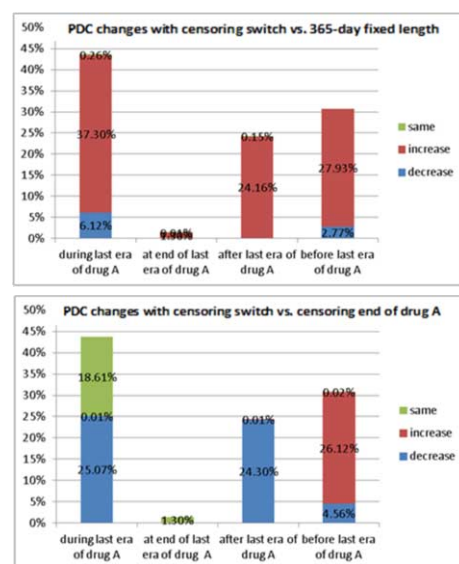


Figure 2- PDC change patterns across three indications by switch index category

Discussion

There are two main findings from this study; 1) drug switching was frequently observed in antihypertensive, antihyperglycemics, and cholesterol lowering drugs in both a commercial claims database and a Medicaid supplement database; 2) censoring drug switching significantly impacts medication adherence measures and classifications in commonly used chronic oral medications. Our study results showed that a fair number of patients switched to another ingredient within one year: near 20% for antihypertensive, 20% for antihyperglycemics, and 6-7% for cholesterol lowering drugs.

These numbers were consistent with previous findings in the literature. Bernard observed that 6-9% of patients switched to new therapeutic class of antihypertensive medication within 12 months [20]. Dailey, et al. found that near 30-40% of patients with antihyperglycemics monotherapy, modified treatment within 1-year follow up period [21-22]. Simpson et al. reported that 18% of patients with statins monotherapy and uncontrolled lipoprotein cholesterol experienced medication switches or titration within 12 months in a managed care setting [23].

Incorporating censoring of drug switching into the adherence estimation significantly affected medication adherence measures. When PDCswitch and PDCend were compared, a modest, but significant PDC difference in the population mean was observed. Further investigation by paired t-test of PDC at patient level, showed that censoring switching might either increase or decrease PDC with a balanced distribution. As a result, the PDC difference at the patient level might average out in the population mean. However, 30% of patients changed their adherence status, either from adherent to non-adherent or from non-adherent to adherent, which suggested that failing to censor PDC based on drug switching may either over-estimate or underestimate patient adherence status. One possible explanation for why 30% of patients changed their adherence status is that the PDCend values for these patients might be near 0.8, thus a minor change in PDC may cause patient adherence status change. However, we investigated that PDCend values of these patients were broadly, rather than narrowly distributed around the 0.8 cutoff. For example, in Optum, the PDCend median was 0.9 (Q1 0.7, Q3 1.0). In contrast, when comparing PDCswitch with PDC365, mean PDC was markedly increased (more than 0.3) in antihypertensive, antihyperglycemics, and cholesterol lowering drugs. The paired t-test at patient level also suggested that PDC increased more than 0.4 when censoring drug switching, and 60-80% of patients changed adherence status, mainly (95-99%) from non-adherent to adherent.

Medication adherence is an important factor linking treatment effectiveness, patient outcomes, and healthcare economic consequences [24]. Non-adherence to prescribed medication regimens relates to poor disease control, causes higher rates of mortality and morbidity, and increases health care costs [25-28].

Over a decade ago, the WHO stated that an effective adherence intervention might have a greater effect on population health than any other medical treatment improvements [29]. Recently, both the Pharmacy Quality Alliance and the National Quality Forum endorsed the PDC-based adherence measures. Accurately assessing patient medication adherence is essential to optimizing drug therapy, as well as to identify non-adherence patients for targeted

interventions. Our study showed drug switching is frequently observed among patients with chronic conditions, and censoring drug switching significantly changed PDC values: less than 20% of patients remained the same PDC value as the PDCend, and less than 0.5% of patients remained the same PDC value as the PDC365 (Figure 2). Failing to censor drug switching may miscalculate the patient adherence measure and consequently misclassify adherence status. An appropriate approach to measure adherence should take drug switching into account when patients switched to a new drug, their medication taking behaviors for previous drugs should be censored, including both the number of days supply and the observation period. At the same time, adherence for the new drug should be measured, starting at the time point when the patient modified their therapy.

Limitations

This study explored how censoring drug switching affected medication adherence measures in patients with single ingredient chronic oral medications. Certain limitations should be acknowledged. First, we only focused on switching between ingredients. We did not investigate dosage change or changes from monotherapy to combination treatment; therefore the switching rate might be under-estimated. Second, patient dispensing patterns (including drug switching and medication adherence) were only followed up for one year and findings may not represent long-term patterns [30]. Third, we use a 30-day grace period as a general observation window to detect drug switching for all 60 ingredients. In clinical practice, a more specific grace period might be applied for certain drug switching due to side effects or perceived efficacy [31]. Fourth, we assumed that patient dispensing data were complete and we have may missed patient dispensing information if a patient changed their insurance plan. However, the MDCD is a Medicaid supplement database which represented a relative stable population, yet demonstrated similar findings with Optum. Finally, we did not study specific drugs and diseases, results from this study only provided general knowledge on how censoring drug switching impacts medication adherence measures. One interesting finding is that censoring switching mainly increases the PDC in antihypertensive and antihyperglycemics, while decreases PDC in cholesterol lowering drugs. A more disease-specified study should be conducted to understand more about the different effects of censoring drug switching on medication adherence measures.

Conclusion

To our knowledge, this is the first study to explore how censoring drug switching affects adherence measures and classifications in common chronically used oral medications. We found that drug switching frequently occurred in real-world practices and significantly affected PDC values in various ways. More importantly, censoring drug switching largely impacted patient adherence status. It is critical to account for drug switching in medication adherence measures in order to accurately assess patient medication taking behavior.

Acknowledgements

This research was partially supported by the Observational Health Data Sciences and Informatics (OHDSI) and Johnson & Johnson, Pharmaceutical (J&J).

References

- [1] K. Bally, R.R. Buechel, P. Buser, P. Tschudia, B. Martinaa, and A. Zeller, Discontinuation of secondary prevention medication after myocardial infarction - the role of general practitioners and patients, *Swiss Med Wkly* **143** (2013), w13896.
- [2] S.J. Bocuzzi, J. Wogen, J. Fox, J.C.Y. Sung, A.B. Shah, and J. Kim, Utilization of oral hypoglycemic agents in a drug-insured US population, *Diabetes Care* **24** (2001), 1411-1415.
- [3] Q. Zhang, C. Zhao, M.J. Davies, L. Radican, and T. Seck, Compliance and persistence with concomitant statin and oral antihyperglycemic therapy, *Am J Manag Care* **17** (2011), 746-752.
- [4] R.P. Hertz, A.N. Unger, and M.B. Lustik, Adherence with pharmacotherapy for type 2 diabetes: a retrospective cohort study of adults with employer-sponsored health insurance, *Clinical Therapeutics* **27** (2005), 1064-1073.
- [5] J. Yeaw, J.S. Benner, J.G. Walt, S. Sian, and D.B. Smith, Comparing adherence and persistence across 6 chronic medication classes, *J Manag Care Pharm* **15** (2009), 728-740.
- [6] G.L. Colombo, E. Rossi, M. De Rosa, D. Benedetto, and A.V. Gaddi, Antidiabetic therapy in real practice: indicators for adherence and treatment cost, *Patient Prefer Adherence* **6** (2012), 653-661.
- [7] A.O. Adeyemi, K.L. Rascati, K.A. Lawson, and S.A. Strassels, Adherence to oral antidiabetic medications in the pediatric population with type 2 diabetes: a retrospective database analysis, *Clinical Therapeutics* **34** (2012), 712-719.
- [8] T. Mabotuwana, J. Warren, J. Harrison, and T. Kenealy, What can primary care prescribing data tell us about individual adherence to long-term medication?-comparison to pharmacy dispensing data, *Pharmacoepidemiology and Drug Safety* **18** (2009), 956-964.
- [9] N.J. Goswami, M. Dekoven, A. Kuznik, J. Mardekian, M.R. Krukas, L.Z. Liu, P. Bailey, C. Deitrick, and J. Vincent, Impact of an integrated intervention program on atorvastatin adherence: a randomized controlled trial, *Int J Gen Med* **6** (2013), 647-655.
- [10] A. Furman, J.L. Meier, R.A. Malmstrom, J.R. Lopez, and S. Schaefer, Comparative efficacy of ezetimibe/simvastatin, rosuvastatin, and atorvastatin in uncontrolled hyperlipidemia patients, *Am J Manag Care* **17** (2011), 538-544.
- [11] J.C. Barner, Adherence to oral antidiabetic agents with pioglitazone and metformin: comparison of fixed-dose combination therapy with monotherapy and loose-dose combination therapy, *Clinical Therapeutics* **33** (2011), 1281-1288.
- [12] A. Adeyemi, K. Rascati, and S. Strassels, The need for oral antidiabetic treatment adherence studies in the pediatric population, *Clinical Therapeutics* **32** (2010), 1306-1307.
- [13] J.P. Swindle, J. Potash, M. Kulakodlu, A. Kuznik, and A. Buikema, Drug utilization patterns and cardiovascular outcomes in elderly patients newly initiated on atorvastatin or simvastatin, *Am J Geriatr Pharmacother* **9** (2011), 471-482.
- [14] R.A. Hansen, J.F. Farley, M. Droege, and M.L. Maciejewski, A retrospective cohort study of economic outcomes and adherence to monotherapy with metformin, pioglitazone, or a sulfonylurea among patients with type 2 diabetes mellitus in the United States from 2003 to 2005, *Clinical Therapeutics* **32** (2010), 1308-1319.
- [15] G. Corrao, S.A. Romio, A. Zambon, L. Merlino, E. Bosi, and M. Scavini, Multiple outcomes associated with the use of metformin and sulphonylureas in type 2 diabetes: a population-based cohort study in Italy, *European Journal of Clinical Pharmacology* **67** (2011), 289-299.
- [16] B.J. Quilliam, A.B. Ozbay, B.E. Sill, and S.J. Kogut, The association between adherence to oral anti-diabetic drugs and hypoglycaemia in persons with Type 2 diabetes, *Diabetic Medicine* **30** (2013), 1305-1313.
- [17] A.V. Gaddi, D. Benedetto, F. Capello, C. Di Pietro, E. Cinconze, E. Rossi, V. De Sando, M. Cevenini, and G. D'Alo, Oral antidiabetic therapy in a large Italian sample: drug supply and compliance for different therapeutic regimens, *Public Health* **128** (2014), 70-76.
- [18] J.M. Overhage, P.B. Ryan, C.G. Reich, A.G. Hartzema, and P.E. Stang, Validation of a common data model for active safety surveillance research, *J Am Med Inform Assoc* **19** (2012), 54-60.
- [19] N.C.f.Q. Assurance, Pharmacy Quality Alliance (PQC) Demonstration Project, in, 2008.
- [20] B.S. Bloom, Continuation of initial antihypertensive medication after 1 year of therapy, *Clinical Therapeutics* **20** (1998), 671-681.
- [21] G. Dailey, M.S. Kim, and J.F. Lian, Patient compliance and persistence with antihyperglycemic drug regimens: evaluation of a medicaid patient population with type 2 diabetes mellitus, *Clinical Therapeutics* **23** (2001), 1311-1320.
- [22] G. Dailey, M.S. Kim, and J.F. Lian, Patient compliance and persistence with anti-hyperglycemic therapy: evaluation of a population of type 2 diabetic patients, *J Int Med Res* **30** (2002), 71-79.
- [23] R.J. Simpson, K. Tunceli, D.R. Ramey, D.R. Neff, D.M. Kern, H.M. Hsieh, D.A. Wertz, J.J. Stephenson, E. Marrett, J.E. Tomassini, and T.A. Jacobson, Treatment pattern changes in high-risk patients newly initiated on statin monotherapy in a managed care setting, *Journal of Clinical Lipidology* **7** (2013), 399-407.
- [24] L. Osterberg and T. Blaschke, Adherence to medication, *N Engl J Med* **353** (2005), 487-497.
- [25] P.M. Ho, J.S. Rumsfeld, F.A. Masoudi, D.L. McClure, M.E. Plomondon, J.F. Steiner, and D.J. Magid, Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus, *Archives of Internal Medicine* **166** (2006), 1836-1841.
- [26] N. Muszbek, D. Brixner, A. Benedict, A. Keskinaslan, and Z.M. Khan, The economic consequences of noncompliance in cardiovascular disease and related conditions: a literature review, *Int J Clin Pract* **62** (2008), 338-351.
- [27] V.J. Zhu, W. Tu, D.G. Marrero, M.B. Rosenman, and J.M. Overhage, Race and medication adherence and glycemic control: findings from an operational health information exchange, *AMIA Annu Symp Proc* **2011** (2011), 1649-1657.
- [28] T.W. Zhu VJ, Rosenman MB, Overhage JM, Facilitating Clinical Research through the Health Information Exchange: Lipid Control as an Example, in: *AMIA Annu Symp Proc*, 2010, pp. 947-951.
- [29] E. Sabaté, *Adherence to Long-term therapies: Evidence for action*, 2003.
- [30] J.S. Benner, R.J. Glynn, H. Mogun, P.J. Neumann, M.C. Weinstein, and J. Avorn, Long-term persistence in use of statin therapy in elderly patients, *Jama-Journal of the American Medical Association* **288** (2002), 455-461.
- [31] J.B. Layton, M.A. Brookhart, M. Jonsson Funk, R.J. Simpson, Jr., V. Pate, T. Sturmer, and A.V. Kshirsagar, Acute kidney injury in statin initiators, *Pharmacoepidemiol Drug Saf* **22** (2013), 1061-1070.

Address for correspondence

Vivienne J. Zhu: zhuv@musc.edu

POSTERS

This page intentionally left blank

I. Connected and Digital Health

This page intentionally left blank

The article ‘SMS Education for the Promotion of Diabetes Self-Management in Low & Middle Income Countries: A Randomized Controlled Trial in Egypt’ has been withdrawn by the authors, Haitham Abaza, Michael Marschollek and Mareike Schulze.

A Study on Saudi Diabetic Patients' Readiness to Use Mobile Health

Hanan Alenazi^a, Mariam Alghamdi^a, Sarah Alradhi^a, Mowafa Househ^b, Nasriah Zakaria^a

^a College of Medicine, King Saud University, Riyadh, Saudi Arabia

^b College of Public Health and Health Informatics, King Saud Bin Abdulaziz University for Health Sciences, Ministry of National Guard Health Affairs, Riyadh, Saudi Arabia

Abstract

The acceptance and feasibility for using mobile health technologies for diabetes management in Saudi Arabia remains to be explored. We created the Mobile Health Effectiveness and Readiness Questionnaire (MHERQ) to measure its validity and reliability in assessing the readiness of Saudi diabetic patients use of mobile health technologies for diabetes self-management. The study was able to measure the validity and reliability of MHERQ and found that MHERQ provides promising results for Saudi diabetic patients' use of mobile health technologies.

Keywords:

Mobile health; Diabetes; Saudi Arabia

Introduction

Saudi Arabia has the second highest diabetes prevalence rate in the Middle East and seventh in the world [1]. The use of mobile technologies in health care, and especially, for chronic disease management, such as Diabetes, has been shown to have positive impacts on health [2]. The purpose of this study was to validate a questionnaire for helping assess the attitude and acceptance of mobile health technologies by diabetics for self-management in Saudi Arabia.

Methods

We created the Mobile Health Effect and Readiness Questionnaire (MHERQ) that was adapted from the Telemonitoring Readiness and Effectiveness Questionnaire (THREQ) [3]. The questionnaire was translated into Arabic following the process of translation recommended by the World Health Organization (WHO) and consisted of three main sections: demographic, clinical data and thirteen questions scored on a five-point Likert scale that related to the readiness in using mobile technologies for diabetes self-management. In March 2017, to validate the questionnaire, a pre-test was conducted on a target population consisting of ten patients visiting the primary care clinic at King Khalid University Hospital (KKUH), in Riyadh, Saudi Arabia. Data were analyzed and the items that did not show acceptable reliability were modified. For the pilot test, a sample size of 30 diabetic patients from KKUH were collected by Face-to-face interviews.

Results

Overall, the MEHRQ showed excellent internal consistency ($\alpha=0.9$). Results also showed that 19 (63%) participants kept their blood glucose values on paper. None of the participants recorded their blood glucose values on a mobile health device or their personal computer.

Participants' response to the thirteen MHERQ items are summarized as follows: 77% of the participants agreed that sending their diabetic data using technology helps them cope better with their diabetes and makes them feel healthier. Also, 80% of the participants agreed that knowing a healthcare professional and having them review their diabetic data gives them a more secure feeling that their condition is being managed.

Results also showed that 63% of the participants were willing to use mobile health from home to transfer blood glucose values to their diabetes professional. 80% of the participants reported that they were willing to ask their healthcare professional questions using mobile technologies.

Other results showed that 60% of the participants were willing to use mobile technologies during holidays to transfer blood glucose values to their diabetes healthcare professional and 70% reported their willingness to ask their healthcare professional questions using mobile technologies while on vacation. 63% of the participants were willing to ask other diabetic patients questions using mobile health technologies, however, 43% of the participants were not willing to use mobile health technologies to transfer blood glucose values to another diabetic patient.

Conclusions

Preliminary results show that MEHRQ to be satisfactory for assessing Saudi patient acceptance for using mobile health technologies for diabetes self-management.

References

- [1] A. Alwin Robert, M. Abdulaziz Al Dawish, R. Braham, M. Ali Musallam, A. Abdullah Al Hayek, and N. Hazza Al Kahtany, Type 2 Diabetes Mellitus in Saudi Arabia: Major Challenges and Possible Solutions, *Current Diabetes Reviews* **13** (2017), 59-64.
- [2] A. Beratarrechea, A.G. Lee, J.M. Willner, E. Jahangir, A. Ciapponi, and A. Rubinstein, The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review, *Telemedicine and e-Health* **20** (2014), 75-82.
- [3] H.E. Buysse, P. Coorevits, G. Van Maele, A. Hutse, J. Kaufman, J. Ruige, and G.J. De Moor, Introducing telemonitoring for diabetic patients: development of a telemonitoring 'Health Effect and Readiness' Questionnaire, *international journal of medical informatics* **79** (2010), 576-584.

Address for correspondence

Corresponding Author: Hanan Alenazi

E-mail address: HanAAlenazi@KSU.EDU.SA

Saudi Diabetic Patients' Attitudes Towards Patient Portal Use and Their Perceived E-Health Literacy

Rawan AlOthman, Nasriah Zakaria, Ahmed AlBarrak

Medical Informatics and E-learning Unit, Medical Education Department,
College of Medicine, King Saud University, Riyadh, Saudi Arabia

Abstract

A cross-sectional study ($n=130$) was conducted to identify diabetic patient's attitudes, patient portal feature preferences, and e-health literacy at two tertiary hospitals in Riyadh, Saudi Arabia. 51% of patients had positive attitude towards the patient portal, 82% think a portal could help with self-care management. 53% of patients had low e-health literacy levels and 9% had high e-health literacy levels. The majority of the study sample have a positive attitude with low e-health literacy levels.

Keywords:

Diabetes Mellitus, Health Literacy

Introduction

The patient portal is a secure platform that allows users to gain access to personal health information regardless of their location. Evidently, such a platform can provide the patient with an expanded potential to improve their process of managing chronic diseases such as diabetes mellitus.

Objective: To assess diabetic patients attitude towards patient portal use for self-care management, preferences towards patient portal features and their general e-health literacy level using a cross-sectional study and convenience-sampling technique.

Methods

Data was collected via self-administrated structured questionnaires [1,2] from diabetic patients who visited outpatient clinics at King Khalid University Hospital (KKUH) and King Faisal Specialist Hospital and Research Center (KFSH&RC) in Riyadh, Saudi Arabia, from February to April 2017.

Results

A total of 130 diabetic patients type 1 and 2, male were 47% and female were 53%. The highest age group was 55-64 years 32%. Only 14.7% had used the patient portal, yet the majority of participants had never used a portal 85%. However, 51% of patients had positive attitude towards the patient portal and 82% think the portal could help with the self-care management. It is found that, 53% of diabetic patients had low e-health literacy level. Whereas, 36% of participant had moderate level along with only 9% who had high e-health literacy level. Over 50% of patients expressed that they prefer to get health information from the Internet. Whilst, 30% had difficulty in understanding and reading health information on

the Internet. The reminder function, scheduling clinic visit and more information on the medication were the most preferred features in a portal.

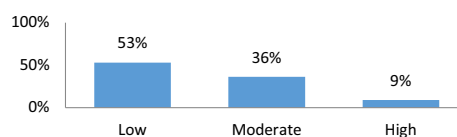


Figure 1 - Diabetic Patients E-health Literacy Levels

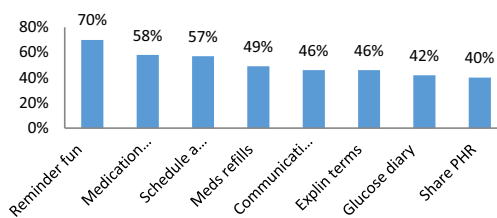


Figure 2 - Diabetic patients preferences of portal features

Conclusion

The majority of diabetic patients had positive attitude along with low e-health literacy levels. The reminder function and scheduling clinic visit were the most preferred features.

Acknowledgements

Great thanks to Abdullah Alqarni for his kind help with the data collection.

References

- [1] Zide M, Caswell K, Peterson E, Aberle DR, Bui AAT, Corey W. Consumers' Patient Portal Preferences and Health Literacy: A Survey Using Crowdsourcing Corresponding Author: 2016.
- [2] Zakaria N, AlBarrak A, Al Fakhri O, Arab N, Matbuli A, Alzahrani A, et al. Development of Saudi e-health Literacy scale (SeHL) for Chronic Diseases in Saudi Arabia: Using Integrated Health Literacy Dimensions. Int J Qual Heal Care. 2017;1-13. In press.

Address for correspondence

Rawan AlOthman

Email: Alothmanrawan@hotmail.com

Development and Deployment of an e-Health System in UNRWA Healthcare Centers (HCs): The Experience and Evidence

Ghada Ballout¹, Najeeb Al-Shorbaji², Amin Shishtawi¹, Fathia Abuzabaida¹, Yousef Shahin¹, Ali Khader¹, Wafa'a Zeidan¹, Yassir Turki¹, Ahmad Al-Natour¹, Rawan Saadeh¹, Majed Hababeh¹, Akihiro Seita¹

¹Department of Health, UNRWA, HQ; ²Director, Knowledge, Ethics and Research (Retired), World Health Organization, Geneva

Abstract

UNRWA recognized the need to develop and introduce eHealth programme based on needs assessment. The web-based e-Health application was developed and deployed in 116 out of 143 HCs in the five fields of operation. Evaluation of the application revealed positive outcomes in terms of reporting requirements, statistics, time saving, efficiency, drug compliance and workflow in general.

Keywords:

Efficiency, Internet, Telemedicine

Introduction

United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) is the main provider of health services to more than 5.8 million Palestinian refugees in its 5 fields of operation: (Jordan, Syria, Lebanon, West Bank and Gaza). In 2009 UNRWA started developing an Electronic Medical Record system. The primary aim of developing the system has been to improve the quality of services, and facilitate patient's flow within the UNRWA HCs.

The UNRWA eHealth system

The UNRWA e-health system is a locally developed web-based, patient-centered system that caters for all types of illnesses. By December 2016, 116 out of 143 HCs, covering almost 90% of the refugee populations, had the e-health system fully functional, and managing 8 million patient visits a year making 90 HCs paperless.

Evaluation of the UNRWA eHealth system

The system was formally evaluated by a group of experts and staff to identify the benefits gained and lessons learned since its deployment in 2011 to end of 2015. The following key findings were identified:

- The daily medical consultations per doctor were reduced on average from 104 to 86 a day;
- The time needed to collect prescribed medication was reduced to 3 minutes;
- The antibiotics prescription rate was decrease from an average of 27.0% to 24.7%;
- The dashboard enabled managers and supervisors to remotely monitor all HCs daily operations and health care provided to patients;
- Comprehensive reporting and statistics on a daily basis;

- 89% of the physicians who were surveyed expressed satisfaction on e-Health, particularly on the fact of time-saving that allowed them to provide better and more attention to patients;
- Managing the crowds in a timely manner and with fairness.

Conclusions

The e-health system is efficient in managing time and service quality, and supports cost containment and client satisfaction.

References

- [1] HealthIT.gov. What are the advantages of electronic health records? <https://www.healthit.gov/providers-professionals/faqs/what-are-advantages-electronic-health-records> (Accessed on 10/09/2012)
- [2] UNRWA. What we do: Health. <https://www.unrwa.org/what-we-do/health> (accessed on 18/06/2016)
- [3] UNRWA Health Department. E-Health (Electronic Medical Records). Annual report 2015.-Amman: UNRWA, 2016. P.13
- [4] UNRWA. UNRWA's new person-centred health service "excellent", say patients. <https://www.unrwa.org/news-room/features/unrwa%E2%80%99s-new-person-centred-health-service-%E2%80%99Cexcellent%E2%80%9D-say-patients> (Accessed on 20 March 2014)
- [5] Dominguez-Gonzalez, J. L. , et. al. The e-Health evaluation report.- Amman: UNRWA, 2016.

Address for correspondence

Najeeb Al-Shorbaji, PhD

Director, Knowledge, Ethics and Research (Rteried), WHO/HQ

Consultant, Knowledge management and eHealth, P. O. Box 524006, Amman, 11934, Jordan

shorbajin@gmail.com shorbajin@plri.de

Cultural Problems Associated with the Implementation of eHealth

Mandirola Briex H-F.^{a,d}, Benitez S.^a, Otero C.^a, Luna D.^a, Masud J.H.B.^b, Marcelo A.^c,
Househ M.^f, Hullin C.^h, Villalba C.^g, Indarte S.^e, Guillen S.^d, Otero P.^a,
Campos F.^{a,d}, Baum A.^a, Gonzalez Bernaldo de Quirós F.^a

^a Hospital Italiano (HIBA), Buenos Aires, Argentina

^b Public Health Informatics Foundation (PHIF), Bangladesh

^c University of the Philippines, Manila, Philippines

^d HL7, Argentina

^e UEIIDISS, Montevideo, Uruguay

^f King Saud Bin Abdulaziz University for Health Sciences, College of Public Health and Health Informatics, Ministry of National Guard Health Affairs, Riyadh, Kingdom of Saudi Arabia

^g Facultad Politécnica, Universidad Nacional de Asunción, Paraguay

^h UC Pontificia Universidad Católica de Chile, Santiago de Chile, Chile

Abstract

A major hurdle in eHealth implementation is that it is difficult to put into practice. In this study, the primary aim was to identify the main barrier associated with implementing eHealth. This study surveyed IMIA members from May to November 2015. From the results, it is clear that medical professionals were recording most of their data by hand. This paper culture is a paradigm that is difficult to break. Cultural factors are the primary barrier in eHealth implementation.

Keywords:

Computer systems; Telemedicine; Medical Informatics

Introduction

Working without paper is not an easy goal because barriers are not solely technological and technology is more accessible. Rather, workflow processes require rethinking and breaking also with aspects of the culture of using paper. [1] eHealth refers to the application of information and communication technologies (ICT) across the whole range of functions that affect health. [2]

Methods

In order to collect data about cultural problems associated with the implementation of eHealth, we conducted a survey. Inclusion criteria for participants included medical informatics experts who are members of IMIA working group forums. The survey was distributed by e-mail using Survey Monkey. The survey included questions that asked about the topics below.

Do you think that it is important:

- To break the culture of paper first?
- To keep in mind local languages and terms used in each region or country?
- To study first the idiosyncrasies of each place and to adapt technology?

Results

A total of 111 responses were received (response rate 92%), where, 27 came from South America, 24 from North America,

27 from Europe, 3 from Africa and 30 from Asia. Responses were excluded if the participant did not answer >2 questions.

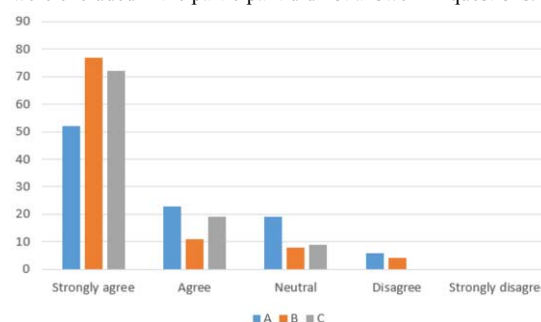


Figure 1 – Responses to survey questions

Conclusion

Cultural barriers are one of the most important barriers in implementing eHealth, [1-2] a finding supported by this study.

References

- [1] N.M. Lorenzi, R.T. Riley. Managing Change. *J Am Med Inform Assoc.* 7 (2000) 2:116–24.
- [2] Digital Agenda for Europe - European Commission. Digital Agenda for Europe. Available from: <http://ec.europa.eu/digital-agenda/> Last accessed 23 November 2015.

Address for correspondence

Dr. Humberto F. Mandirola Briex hmandirola@gmail.com

Management of In-field Patient Tracking and Triage by Using Near-Field Communication in Mass Casualty Incidents

Po-Liang Cheng^{ab}, Yung-Cheng Su^a, Chung-hung Hou^a, Po-Lun Chang^b

^a Emergency Department, Dalin Tzu Chi Hospital, Chiayi County, Taiwan, China

^b Biomedical Institute, National Yang-Ming University, Taipei City, Taiwan, China

Abstract

Near field communications (NFC) is an emerging technology that may potentially assist with disaster management. A smartphone-based app was designed to help track patient flow in real time. A table-drill was held as a brief evaluation and it showed significant improvement in both efficacy and accuracy of patient management. It is feasible to use NFC-embedded smartphones to clarify the ambiguous and chaotic patient flow in a mass casualty incident.

Keywords:

Disasters; Smartphone; Triage.

Introduction

Due to the limited information in the early phase of rescue operation in the mass casualty incident (MCI), the government, emergency medicine service (EMS) systems and the population in Taiwan usually get the ambiguous information of damage. Utilization of the medical electronics with well-established infrastructure of information communication technology to manage affected population in such scenario is thought to be practical and useful. Near field communications (NFC) is an emerging technology that may be potential to develop novel disaster management systems [1]. Recently, an NFC based medication system may be used to effectively reduce medication errors in a simulated environment [2]. The aim of the project is managing the dataflow of the patients with the NFC-embedded smartphones.

Methods

The information platform as an independent smartphone app was programmed by Dr. Cheng under the developed environment of Android Studio 2.0. A unique NFC tag would be placed on the patient in the incident scene by the emergency medical technician (EMT). The EMT would operate the smartphone-based app and the patient information would be transmitted into the encrypted NFC tag. Subsequently, the NFC tag would be transferred along with the patient so that the location and condition would be updated every time after the NFC induction. Instant and editable information will be provided to the clients including the EMTs in the incident scene, the staffs in the hospitals and the commanders. All data in transit will be encrypted because of the security concerns. Afterward, a table-drill was held to evaluate the efficacy and efficiency. There were 50 EMT trainees recruited and then divided into 2 similar groups. A 30-minutes mini-lecture about

simple triage and rapid treatment (START) was instructed by Dr. Hou in advance. Fifty scenario cards were given to the two different groups. In the first group, they were asked to complete triage by using the NFC-embedded smartphone. In the second group, they would complete other triage by using traditional paperwork alone. Finally, they must calculate and report the total injuries and their dispositions. The time and accuracy was recorded by comparison.

Results

There was significant difference in the total triage time between the NFC group (32.38 minutes) and the paperwork group. (49.58 minutes) ($p < 0.001$). Besides, it was significant difference in sorting the information between the NFC group (29 seconds) and the paperwork group. (8.08 minutes) ($p < 0.001$). It means that it is time-consuming to find out which destination and the transfer time. The accuracy was also significant difference between the NFC group (49/50, 100%) and the paperwork group (31/50, 96%).

Conclusion

It is feasible to use a NFC-embedded smartphone to triage and tracking the patient dispositions in the early phase of a mass casualty incident. The limitation in this brief study are the small convenient samples and scenarios.

Acknowledgements

Funding by the Ministry of Science and Technology, Taiwan.

References

- [1] Bellini P et al. Mobile emergency, an emergency support system for hospitals in mobile devices: pilot study, *JMIR Res Protoc* **23**(2) (2013), e19.
- [2] Bachmann DJ et al. Emergency Preparedness and Disaster Response: There's An App for That, *Prehosp Disaster Med* **30**(5) (2015), 486-90.

Address for correspondence:

Corresponding author: Po-Liang, Cheng. Email: kupffer.orz@gmail.com. Phone: +886-5-2846000 ext. 5838

Development of Day Care Management Systems for Elderly

Ming-Chu Chiang^a, Yu-Ting Yeh^{a,b}, Chien-Tsai Liu^a

^a Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei, Taiwan, China,

^b Taipei Medical University Shuang-Ho Hospital, Taipei, Taiwan, China

Abstract

The purpose of this study is to build a practical, highly efficient management information system (MIS) for elderly day care (EDC) centers to improve elders' lives. Based on system requirements and the management operation guidelines for EDC, we have developed an MIS. The introduction of the system will help EDC management to be more accurate, comprehensive, and efficient.

Keywords:

Management Information Systems; Adult Day Care Centers; Long-Term Care.

Introduction

Aging population is a serious challenge that many developed countries have to face together in the twentieth-first century, and so is the case in Taiwan. The elders (of age 65 and over) accounted for 7% of Taiwan's population in 1993, entering an aging society. According to a projection by Taiwan's National Development Council of the Executive Yuan, Taiwan will become an aged society in 2018 (increasing to 14%), and further to a super-aged society in 2026 (up to 20%)[1,2]. Taiwan started to promote the "Ten-Year LTC Plan" (2008-2017). The plan is to expand and popularize the capacity of the resources of the service system, strengthening the popularization and the localization of the long-term care services, in order to raising the quality of the long-term caring.

Methods

This study is in the cloud platform to build EDC center system, improve the EDC center of the manual management process, and develop a series of mobile integration system in future. Microsoft cloud development tools are leveraged to integrate automated physiological measurement equipment, formulate form design standards to deal with different management forms, this set of EDC center intelligent management functions, including authorization management, case filing, case management etc. [Figure 1]. Therefore, this study provides the development of science and technology management of community care services, which is conducive to the country to continue to promote long-term care services to enhance the quality and achieve good results in community service.

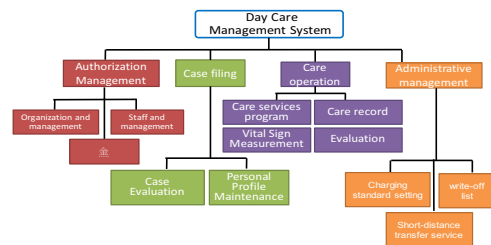


Figure 1– Structure of Day Care Management System

Results

The following is the result of building a day care center management system. The system is developed on Microsoft ASP.NET MVC 5 Framework and built on the MS Azure cloud platform. Moreover, the software design is based on the Model-View-Controller structure and use the database of MS SQL server. The system develops many functions to improve the daily operations. It will prompt the staff to do the regular evaluation, for example ADL (activities of daily living), iADL (Instrumental activities of daily living). Those evaluation records will be kept in the system for future inquiry. The evaluation sheet includes the profile information, the disabilities and the score of the case. The system also provides foolproof mechanism and avoids the data repetition to confirm the data accuracy. Each month will produce list, lead to the Government subsidy. The system win lots of praise from the user and the care staff during the trial phase. The system is able to shorten the procedure of operation.

Conclusion

The system won lots of praise from the caring staff during the test runs, e.g., when the vital signs measurement is completed, the system simultaneously marks the elders' presence records and simplifies the operation procedures. The prompting of the daily pending tasks such as conducting the periodic evaluation, doctor consultation, medication, and etc., has promoted the caring quality.

References

- [1] National Development Council (2016) the Population of the Republic of China (2016-2061 Year)
- [2] Ministry of the Interior (2016) Monthly Bulletin of Statistics, Ministry of the Interior 1.11 Data The age of single - age population and the median age of the population.

Address for Correspondence

Chien-Tsai Liu 250 Wu-Xing Street, Taipei, Taiwan; Email: ctiu@tmu.edu.tw

Use of Person-Generated Health Data in Kinect-Based Stroke Rehabilitation Systems: A Systematic Review

Gerardo Luis Dimaguila, Kathleen Gray, Mark Merolli

Health and Biomedical Informatics Centre, University of Melbourne, Melbourne, Victoria, Australia

Abstract

It is important for patients to have access to the health information they generate, for them to be actively engaged in their own healthcare. This is relevant to Kinect-based post-stroke rehabilitation systems; as such there is a need to review the literature based on person-generated management and utilisation. Previous systematic reviews on Kinect for stroke rehabilitation have not used this as part of their criteria. This systematic review fills that gap.

Keywords:

Review, Systematic; Stroke; Health Technology

Introduction

Person-generated health data (PGHD) are health, wellness and clinical data that people generate, record and analyse for themselves [1]. While PGHD's importance for individuals or patients have yet to be conceptually defined, it is well known that when patients understand their illness, they become active problem solvers and improve their health behaviour, e.g., people will stop smoking when they personally see the connection between that and an illness they are experiencing [2]. It is therefore important for patients to have access to the health information they generate. This is relevant to post-stroke rehabilitation systems using body-tracking technology Kinect, which have been developed as a response to the need for the effective home-based rehabilitation that requires less professional and financial resources [3-5] because such systems generate patient-relevant data. There is therefore a need to review the literature on Kinect-based stroke rehabilitation to understand if and how person-generated data are managed and utilised. Previous systematic reviews [3-5] have focused on describing the intervention, methodologies and results, and have not given attention to person-generated data management.

Methods

The terms "Kinect", "stroke", "cerebrovascular accident", "CVA" and "rehabilitat*" (% for ACM) were used to search for peer-reviewed English articles in PubMed, BioMed Central, IEEE Xplore, and ACM. There was no date restriction. Papers included are those that used Kinect as a body-tracking device for a technology-based stroke rehabilitation system. The search was last updated on December 7, 2016.

Results

A total of 35 papers were included in the review, out of 88 search results. It was discovered that patient data access came in the form of feedbacks. 15 of the studies provided visual or

auditory feedbacks for patients to follow correct gestures or positions; 3 studies provided task counts, e.g., completed/to be completed; 5 studies simply provided game scores, not clearly related to performance; and 13 studies did not discuss provision of feedback or data utilisation. The numbers include 1 paper that provided both performance feedback and task counts. For the complete list of papers, please contact the corresponding author.

Conclusions

The results show that there is insufficient attention given to person-generated data from Kinect-based stroke rehabilitation systems. While most studies provide some form of feedback, they do not allow the patients to actively engage in their own rehabilitation, nor do the studies try to understand the health behaviour impact of providing data access to patients. This is indicative of the need for future researchers of technology-based rehabilitation to consider PGHD and patient access to information in their systems' design and implementation.

Acknowledgements

The primary author would like to acknowledge the support of the Health and Biomedical Informatics Centre; and Newman College (University of Melbourne), his organisational sponsor.

References

- [1] S.T. Rosenbloom, Person-generated health and wellness data for health care., *Journal of the American Medical Informatics Association* **23** (2016), 438-439.
- [2] M.B. Sagar, Elizabeth, Participatory medicine: model based tools for engaging and empowering the individual., *Interface focus* **6** (2016).
- [3] A.F. Da Gama, Pascal; Teichrieb, Veronica; Navab, Nassir, Motor rehabilitation using Kinect: A systematic review, *Games for Health* **4** (2015), 123-135.
- [4] H.M.K. Hondori, Maryam, A review on technical and clinical impact of Microsoft Kinect on physical therapy and rehabilitation, *Journal of Medical Engineering* (2014).
- [5] D.C. Webster, Ozkan, Systematic review of Kinect applications in elderly care and stroke rehabilitation, *Journal of Neuroengineering and Rehabilitation* **11** (2014).

Address for correspondence

Gerardo Luis Dimaguila

E-mail : dgl@student.unimelb.edu.au

A Serious Game for Anterior Cruciate Ligament Rehabilitation: Software Development Aspects and Game Engagement Assessment

Marcos Cordeiro d'Ornellas^a, Alison Santos Machado^a,
Jefferson Potiguara de Moraes^b, Ana Lúcia Cervi Prado^b

^a Grupo de Computação Aplicada em Saúde (CA+SA), Universidade Federal de Santa Maria (UFSM), Santa Maria-RS, Brasil,

^b Serviço de Fisioterapia, Hospital Universitário de Santa Maria (HUSM), Santa Maria-RS, Brasil

Abstract

This work presents the steps for developing a serious game that allows the interaction through natural gestures, whose main purpose is to contribute to the treatment of individuals who have suffered an injury to the anterior cruciate ligament (ACL). In addition to the serious game development process, the users' gaming experience were performed. Through the evaluation assessment, positive results were obtained in relation to various aspects of the game engagement, proving the playful factor of this activity.

Keywords:

Anterior Cruciate Ligament, Rehabilitation, User-Computer Interface

Introduction

The knee is one of the joints that have a higher incidence of injuries resulting from sports activities. Among the most frequent injuries in the knee joint are the injuries to the anterior cruciate ligament (ACL). Most ACL injuries in soccer players happen when a player lands unadroitly in the wake of bouncing. Since ACL tears do not mend, those wishing to come back from a soccer injury are urged to have the ACL reconstructed [1].

This paper presents a serious game for ACL rehabilitation using natural user interfaces. The game helps the therapist in assessing the quality and efficiency of the treatment and encourages patients using a serious game that uses quantitative measures to assess game engagement while using existing interfaces for user-computer interaction.

Methods

The game (see Fig. 1) has several settings that can be changed by the physical therapist. The simplicity of the controls offers an amount of variation when players are shooting. They can take a few shots to get the direction and spin aspects just right to score.

In order to measure engagement, players were asked to complete the Game Engagement Questionnaire. The questionnaire is a nineteen-item scale that assesses several aspects related to gameplay engagement. These include presence, flow, immersion, and absorption. Rasch analysis of (GEQ) responses and relationships with other questionnaires and with participant behavioral responses provided indications that the game engagement questionnaire is a reliable and valid measure of engagement in playing video games [2].



Figure 1 – Screenshot of the game setting menu on the left side. A small screen on the lower right corner shows an infrared image of the subject while playing the game.

Results

To compare each of the engagement scales between conditions, a statistical analysis was performed using two-way analysis of variance (ANOVA). Based on the results achieved during the in-game engagement assessment, the prototype performed to specification quite well.

Conclusion

Through tests conducted by the GEQ, it was possible to verify that the developed game reached the goals being evaluated positively. Future studies suggest the evaluation of GEQ with individuals who participate in the treatment of recovery after ACL injuries in order to evaluate the engagement of people who have limited dexterity.

References

- [1] K.E. Wilk, L.C. Macrina, E.L. Cain, J.R. Dugas, J.R. Andrews, Recent advances in the 337 rehabilitation of anterior cruciate ligament injuries. *J Orthop Sports Phys Ther.* 2012 338 Mar;42(3), 153-71.
- [2] K. L. Norman, Game engagement questionnaire: A review of two papers, 2013; 25(4),278-283.

Address for correspondence

Dr. Marcos Cordeiro d'Ornellas, marcosdornellas@gmail.com, Av. Roraima, 1000, Prédio 07, Anexo B, Sala 388 UFSM- Campus Universitário 97105-900 Santa Maria-RS – Brasil.

Comparing the Human Papillomavirus Vaccination Opinions Trends from Different Twitter User Groups with a Machine Learning Based System and Semiparametric Nonlinear Regression

Jingcheng Du^{a*}, Jing Huang^{b*}, Rui Duan^b, Yong Chen^b, Cui Tao^a

^a School of Biomedical Informatics, University of Texas Health, Houston, TX, USA,

^b Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, USA

Abstract

HPV vaccination refusal is a serious public health issue. Opinions on Twitter have are influential to potential consumers on vaccination behaviours. Public opinions toward HPV vaccination were extracted from Twitter by leveraging machine learning models. We used semiparametric nonlinear regression models to study the association between people's public opinions on HPV vaccination at Twitter and the number of followers. Trend patterns of opinions from Twitter users groups with different number of followers were identified and compared.

Keywords:

Vaccination; Machine learning, Social media.

Introduction

Vaccination refusal of human papillomavirus (HPV) vaccines is a serious health problem, attracting attention from informatics researchers [1]. Social media information could affect consumer health behavior [2]; analysis of public HPV opinions from social media is very important. Twitter provides a rich data source to access public opinions on HPV vaccines. Before quantifying the influence of Twitter users on others' vaccination behavior change, it is interesting to study the association between the users' own opinions with the number of followers. Successful quantification of such association may help design more effective interventions to mitigate the vaccination refusal.

Methods

Tweets were collected from November 2, 2015 to March 28, 2016 using HPV vaccine related keywords through the Twitter Application programming Interface. We leveraged Support Vector Machine (SVM) models [3] to categorize the tweets into three opinion groups including "Positive", "Negative", and "Neutral". Other tweets were grouped as "Unrelated".

Tweets were group into four subgroups by the posters' followers numbers using quartile function. We then fit polytomous logistic regression models of the categorized opinions on each subgroup for number of followers and calendar time, using polynomial splines.

Results

Among the 184,214 tweets collected, 73,436 tweets were grouped as "Unrelated", 39,704 as "Positive", 35,428 as "Negative", and 35,591 as "Neutral". The number of tweet followers were grouped into four groups: 0-121, 121-441, 441-

1,460, and 1,460 or more. The number of followers was significantly associated with vaccination opinions ($p \leq 0.01$).

During the study period, among all users groups except 121-441, we observed similar trends our polytomous logistic regression models with splines. More specifically, we found that the probability of "Negative" opinions decreased until 02/2016 and increased afterwards; the probability of "Positive" opinions increased and then decreased starting 02/2016. User group trends with 121-441 followers differed from other groups with decreased "Negative" opinions until 02/2016 and increased "Positive" opinions afterwards.

Conclusions

Based on our previous work, we leveraged a machine learning based approach to extract public opinions toward HPV vaccination from Twitter. By grouping users based on the number of followers and fitting logistic regression models, we accessed and compared opinions trends. Except for Twitter users groups with 121-441 followers, the trend patterns were similar. Analysis of the similarity and difference in opinions patterns can be provided to health professionals to more efficiently and precisely lead the the public opinions on Twitter. A larger Tweets corpus is being collected and more robust conclusions will be made based on the larger dataset.

Acknowledgements

The authors gratefully acknowledge the support from the National Library of Medicine of the National Institutes of Health under Award Number R01LM011829, and the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant # RP160015).

References

- [1] A.G. Dunn, J. Leask, K.D. Mandl, E. Coiera, Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study, *J Med Internet Res* 17 (2015), e144.
- [2] 24 Outstanding Statistics & Figures on How Social Media has Impacted the Health Care Industry [Referral MD, edited 2017]. Available from: <https://getreferralmd.com/2013/09/healthcare-social-media-statistics/>.
- [3] J. Du, J. Xu, H. Song, X. Liu and C. Tao, Optimization on machine learning based approaches for sentiment analysis on HPV vaccines related tweets, *J Biomed Semantics*, under review.

Address for correspondence

To whom correspondence should be addressed: cui.tao@uth.tmc.edu.

* These authors contribute equally to the paper.

Social Media as a Communication Support for Persons with Mild Acquired Cognitive Impairment: A Social Network Analysis Study

Aboozar Eghdam^a, Ulrika Hamidi^a, Aniko Bartfai^b, Sabine Koch^a

^a Health Informatics Centre (HIC), Department of Learning, Informatics, Management and Ethics (LIME), Karolinska Institutet, Stockholm, Sweden,

^b Division of Rehabilitation Medicine, Department of Clinical Sciences Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden

Abstract

This study was conducted as a social network analysis of a Facebook group for Swedish speaking persons (1310 members) with perceived brain fatigue after an illness or injury to the brain to address the lack of research examining social media and the potential value of on-line support for persons with mild acquired cognitive impairment.

Keywords:

Social Support, Cognitive Dysfunction, Social Media

Introduction

One of the leading causes of disability in many developed countries is brain injury. Mild Acquired Cognitive Impairment (MACI) is a term used to describe persons with non-progressive Mild Cognitive Impairment (MCI) acquired from a brain injury. They can have multiple cognitive (e.g., memory impairment) and/or somatic (e.g., headaches, fatigue) symptoms in addition to mild physical disabilities and their access to rehabilitation services to improve their affected skills is limited [1]. Social media has the potential to increase social participation and support for the well-being of individuals with chronic medical conditions [2]. To date, Facebook is the most popular social medium for different types of communication. However, there is a lack of knowledge about the potential use of Facebook as a mean of communication for persons with potential MACI.

Objective

The aim of this study was to explore how persons with MACI, specifically persons with perceived brain fatigue after brain injury, communicate through Facebook, to classify the content of the communication and to visualize the frequency and types of interactions.

Methods

A social network analysis (quantitatively and qualitatively for a period of a whole year) of the interactions between members of a public Facebook group for Swedish speaking persons (1310 members) with perceived brain fatigue after an illness or injury to the brain was performed.

Results

The results showed how members use social media technology and Facebook as a means for communication and support for

their condition. Individual group members showed very different patterns of communication and interaction. However, for the group as a whole, the most frequent topics in their communication were related to informational support and banter in posts, and socialization in comments. The findings also showed that the majority of members only communicated with few other members and had few direct communications. The most used communication feature of Facebook was likes in form of “thumbs-up”.

Conclusion

This study indicated that social media and in this case Facebook is used for communication and social support by persons with MACI and revealed that their communication behavior is similar to the healthy population. Further studies relating specific cognitive problems of the participants to the use of social media would provide more reliable results for this specific group.

Acknowledgements

This study was financed by the Health Informatics Centre, Karolinska Institutet and it has been part of in the PhD thesis of the main author [3].

References

- [1] Eghdam A, Scholl J, Bartfai A, Koch S. Information and communication technology to support self-management of patients with mild acquired cognitive impairments: systematic review. *Journal of medical Internet research*. 2012;14(6):e159.
- [2] Kamel Boulos MN, Wheeler S. The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education. *Health information and libraries journal*. 2007;24(1):2-23.
- [3] Eghdam, A. Understanding how persons with mild acquired cognitive impairment use and experience information and communication technology: an exploratory study. *Karolinska Institutet*. 2016.

Address for correspondence

Aboozar Eghdam
Health Informatics Centre, Dept. of Learning, Informatics, Management and Ethics Karolinska Institutet, Tomtebodavägen 18 A - 4th floor, SE- 171 77 Stockholm
Phone: +46 8 524 864 57
E-mail: aboozar.eghdam@ki.se.

Rapid Development and Distribution of Mobile Media-Rich Clinical Practice Guidelines Nationwide in Colombia

José F Flórez-Arango^{a,b}, M Sriram Iyengar^a, Indira T Caicedo^c, German Escobar^c

^a College of Medicine, Biomedical Informatics Group, Texas A&M University, College Station, TX, USA

^b Escuela de Ciencias de la Salud, Universidad Pontificia Bolivariana, Medellín, Antioquia, Colombia

^c Oficina de Calidad, Ministerio de Salud y de la Protección Social, Colombia

Abstract

Development and electronic distribution of Clinical Practice Guidelines production is costly and challenging. This poster presents a rapid method to represent existing guidelines in auditable, computer executable multimedia format. We used a technology that enables a small number of clinicians to, in a short period of time, develop a substantial amount of computer executable guidelines without programming.

Keywords:

Practice Guideline, Multimedia, Telemedicine.

Introduction

Compliance with evidence-based Clinical Practice Guidelines (CPGs) by health workers, and consumers (patients and caregivers) can result in improved patient outcomes [1]. The Ministry of Health in Colombia (MinSalud) has developed almost 60 CPGs across multiple specialties including pediatrics, infectious- and chronic diseases, and cancer. These are published as PDFs. Factors preventing widespread CPG use include guideline complexity, presentation as linear, non-interactive text, lack of multi-media. These make it very difficult for many to understand and comply with CPGs. Another issue is that it is impossible to track the usage of CPGs.

Methods

MinSalud sponsored a project that takes advantage of the widespread adoption of smart-phones (typically Android OS) in Colombia and a technology, originally developed at NASA, called GuideVue®(GV), for developing media-rich interactive guidelines (MRIGs) without programming. These MRIGs are executed on the web (<http://gpcdigital.minsalud.gov.co>) and also on Android smartphones. MRIGs, based on Persuasive Technology [2] and Felder-Silverman Learning [3] theories, have been shown to increase compliance, decrease errors and perceived workload among health workers [4]. After downloading on the smartphone, connectivity is not needed to execute these MRIGs. Usage tracking is one of the features of GV technology. Twelve healthcare professionals and an editorial coordinator converted PDF-based CPGs into MRIGs using GV Author, Windows software that enables MRIG creation, by using a visual, no-coding, point-and-click, drag-and-drop graphical user interface. Although none of these individuals had any training in programming they rapidly became proficient in GV Author. These MRIGs were carefully reviewed by MinSalud prior to release. Although the CPGs varied widely in scope, topic certain heuristics were designed to apply to all of them. These include: i) Each step expressing

a single action, concept or decision, ii) No more than 5 decision choices in each step, iii) Color-coded decision buttons, iv) Still images and videos to assist in the understanding of the step's content; videos only when an action is to be illustrated (e.g., performing exercises).

Results

In early 2016, a total of 116 MRIGs were created by 12 individuals in 8 weeks, covering 53 topics. Of these, 53 MRIGs were targeted to healthcare workers and 63 to consumers. The system was officially launched on 14/oct/16. Guideline changes can easily be propagated by editing the MRIG in the Author visual editor software. The project is sustained by MinSalud to enhance health literacy among both healthcare consumers and professionals.

Discussion

The GV Author software enabled a small number of non-programmers to create 100+ MRIGs in 8 weeks. These individuals were able to rapidly become productive, typically with only 4 hours of training. For almost all 53 topics two versions were created: one for healthcare professionals, the other for consumers. The consumer version explained the same topic in layman's terms.

Conclusion

The technology described has great potential to support rapid development of media-rich clinical practice guidelines on web and smartphone platforms.

References

- [1] G. García-Sáez, M. Rigla, I. Martínez-Sarriegui, E. Shalom, M. Peleg, T. Broens, B. Pons, E. Caballero-Ruiz, E. Gómez and M.E. Hernando, Patient-oriented Computerized Clinical Guidelines for Mobile Decision Support in Gestational Diabetes, *J. Diabetes Sci. Technol.*, **8** (2014), 238–246, 2014.
- [2] B. Fogg, Persuasive technology: using computers to change what we think and do, *Ubiquity*, (2002), 5.
- [3] R. Felder, L. Silverman, and B. Solomon, Index of Learning Styles (ILS), *Skynet.le*, (2000), 1–10.
- [4] J. F. Florez-Arango, M. Sriram Iyengar, K. Dunn, and J. Zhang, Performance factors of mobile rich media job aids for community health workers, *J. Am. Med. Informatics Assoc.*, **18** (2011), 131–137.

Address for correspondence

Jose F Florez-Arango florezarango@medicine.tamhsc.edu

The Diffusion of mHealth Applications

Parisis Gallos, John Mantas

Health Informatics Laboratory, School of Health Sciences, National and Kapodistrian University of Athens, Athens, Attica, Greece

Abstract

The aim of this preliminary study is to present the usage behaviour of mHealth applications. A survey was conducted among 352 Smartphone users; 109 of them had used mHealth applications as well. The usage of these applications seemed to be quite low. Young people had used mHealth applications more than the elders. Also, there was not found a statistically significant relationship between the gender and the usage of mHealth applications.

Keywords:

Telemedicine, Mobile Applications; Surveys and Questionnaires

Introduction

A lot of attempts have been made lately on the development of mHealth applications [1-4] in Greece. Additionally, several studies have been made related to mHealth technologies [5,6]. According to relative surveys, mHealth applications usage has been doubled in the last two years [7]. The aim of this preliminary study is to present the diffusion of mHealth applications' usage among Smartphone users.

Methods

A simple anonymous questionnaire was distributed on-line, over the social media, in order to record the stakeholders' frequency of usage of mHealth applications. The questionnaire was in Greek language and included questions related to the usage of the mHealth applications. The size of the sample was 352 persons who were using Smartphones. The data analysis was conducted using the SPSS.

Results

Only 109 persons had used a mobile health application (31.1% of the total). The 33.9% of those were males and 66.1% were females. The average of the age was 26.54 years old, with Std. Deviation of 6.78 years. The minimum age was 18 and the maximum 58 years old. The median of the age was 25.50 years old. In the statistical hypothesis testing, age found to be related with the mHealth applications usage ($p = 0.03$). On the other hand, there was not a statistically significant relationship between the gender and the usage of mHealth applications ($p = 0.87$).

Based on the study findings, it can be assumed that the majority of the responders do not use any mHealth applications yet. This can be explained by the fact that there are not a lot of mHealth applications available in their language or maybe there are having thoughts about the validity of the applications' content. Also, young people are using mHealth applications more than the elders as they might be more familiar with the mobile

technology. In addition, males seems to use the mHealth applications as much as the females.

Conclusions

According to the results, it can be assumed that mHealth applications' usage has a low diffusion on Smartphone users in Greece. Also, mHealth technology seems to be more popular among the youths. Current research actions include a further investigation of the mHealth technology diffusion, and the effect of this technology in people's health status.

References

- [1] K. Karagiannaki and et al, mMamee: A mHealth Platform for Monitoring and Assessing Maternal Environmental Exposure, *2015 IEEE 28th International Symposium on Computer-Based Medical Systems, Sao Carlos* (2015), 163-168.
- [2] M. Dimitriou and et al., Evaluation of an Electronic Application for Nutritional Information in Food Service Outlets: A Pilot mHealth Application, *International Journal of Reliable and Quality E-Healthcare (IJRQEH)* 6(1) (2017), 46-58.
- [3] P. Gallos and J. Mantas, "SMART Travel Health" Application, *Stud Health Technol Inform* 210 (2015), 1038.
- [4] M. Bitsaki and et al., An Integrated mHealth Solution for Enhancing Patients' Health Online, *6th European Conference of the International Federation for Medical and Biological Engineering* Springer International Publishing (2015), 695-698.
- [5] P. Angelidis and et al., The hackathon model to spur innovation around global mHealth, *Journal of Medical Engineering & Technology* (2016), 1-8.
- [6] P. Gallos and J. Mantas, The "SMART Travel Health" Mobile Application Assessment, *Stud Health Technol Inform* 213 (2015), 227-229.
- [7] Accenture Consulting, *Patients Want a Heavy Dose of Digital* (2016).

Address for correspondence

Parisis Gallos

Papadiamantopoulou 123

Athens, 11527, Greece.

e-mail: parisgallos@yahoo.com

tel: +302107461448.

Warfarin Guide: Co-Design of a Mobile Computer-Assisted Anticoagulant Therapy

Martin Gerdes^a, Santiago Martinez^b, Berglind Smaradottir^a, Rune Fensli^a, Jarle Jortveit^c

^a Department of Information and Communication Technology, University of Agder, Grimstad, Norway

^b Department of Health and Nursing Sciences, University of Agder, Grimstad, Norway

^c Sørlandet Hospital HF, Arendal, Norway

Abstract

Patients with some types of cardiovascular disease are prescribed anticoagulation therapy with Warfarin in order to control the ability of blood clotting. This work presents a co-designed mobile application, called Warfarin Guide, for a computer-assisted anticoagulant therapy. The application addresses the challenges that unexperienced patients may find when having to remember to regularly check their INR values and make temporary adjustments for INR value fluctuations that are not easy to interpret without direct medical advice.

Keywords:

International Normalized Ratio, Computer-Assisted Drug Therapy, Medical Informatics

Introduction

Patients with some cardiovascular disease types, e.g., mechanical heart valves, are prescribed lifelong anticoagulation therapy with Warfarin to prevent thromboembolic complications. The intensity of the anticoagulation therapy is based on the International Normalized Ratio (INR) value. A low INR value indicates a risk of increased blood clotting which can cause life-threatening conditions, while a high value of INR indicates a high risk of bleeding. Food intake (especially vitamin K-rich vegetables), alcohol consumption, physical exercise, illness and stress may also influence the INR values. Patients treated with Warfarin are recommended to self-monitor their INR values with a portable coagulometer device that samples whole blood obtained by fingerprick, and self-management of the Warfarin therapy. It might be difficult, in particular for unexperienced patients, to regularly check their INR values and make necessary changes in warfarin dosing based on the measured values, in light of previous studies which have reported low time within therapeutic INR range.

The aim of this project is to develop an easy-to-use mobile application for computer assisted anticoagulation therapy. The project is divided in two phases: 1) a co-design, 2) a usability test of the application and 3) a randomised control trial. This work presents the results of the first phase.

Methods

The co-design process was carried out with 8 students of the Master of Health and Social Informatics' at the University of Agder (UiA), Norway and iteratively tested in the Usability laboratory at UiA with patients who were prescribed an anticoagulant therapy.

The Master students co-designed a non-functional prototype of the Warfarin Guide application (WGA), and the explanation of their prototype was recorded on camera. The application developers watched the recordings and implemented a working prototype. The prototype was iterated through 2 usability tests with 5 anticoagulant therapy patients. The first test included a cognitive walkthrough process with an individual interview after using the application. The second test targeted specific tasks of the refined user interface.

Results

The co-design process allowed the Master students to create a non-functional prototype of the application, including different features based on their experience of working with patients prescribed with an anticoagulant therapy. The co-design provided developers with a useful insight about what was important for end-users, prioritising meaningfulness of messages and user friendliness of the user interface (UI). The cognitive walkthrough was useful to understand, step by step, how patients interacted with the application in order to obtain the exact dose, keep track of their INR values and set up a reminder for next dosage intake. The usability tests helped to iteratively improve the UI. The individual interviews allowed the research team to understand the way that the WGA could address users' needs.

Conclusions

Self-management medication for anticoagulation therapy presents two challenges to unexperienced patients: remember and maintain a regular monitoring of their INR values; and calculate an exact dosage intake adjusted to the nonuniform fluctuations of INR values. The co-designed WGA tested in the laboratory provided an up-to-date monitoring of INR values and an individualised calculation of dosage adjustment, including a valuable reminder function for medication intake and upcoming measurements. The adherence, effectiveness and benefits of the computer-assisted anticoagulant therapy will be clinically verified in a clinical randomised control trial.

Acknowledgments

The authors would like to thank all the anticoagulant therapy patients who participated without bias in the project.

Address for correspondence

Martin Gerdes: martin.gerdes@uia.no

Patient-Clinician Co-Design Co-Participation in Design of an App for Rheumatoid Arthritis Management via Telehealth Yields an App with High Usability and Acceptance

Rebecca Grainger^{a, b}, Hermaleigh Townsley^a, Tobias Langlotz^c, William Taylor^{a, b}

^a Department of Medicine, University of Otago Wellington, Wellington, New Zealand,

^b Hutt Hospital, Hutt Valley District Health Board, Lower Hutt, New Zealand

^c Department of Information Science, University of Otago, Dunedin, New Zealand

Abstract

People with chronic conditions like rheumatoid arthritis (RA) self-manage on a day to day basis. They may be able to assess disease activity and communicate this via an app to their healthcare team to enable clinical review for medical management at the most appropriate times. This work describes the successful co-design of a patient-held app for monitoring and communication of RA disease activity.

Keywords:

Telehealth, Patient participation, Arthritis, rheumatoid.

Introduction

Traditional management of rheumatoid arthritis (RA) is by rheumatologist review 3 to 6 monthly. This is not sustainable in many healthcare systems as demand exceeds rheumatologist capacity, with these demands projected to increase. Mobile software (an application or app) could enable patients to provide remote monitoring of health status via patient reported outcomes (PROs), patient-performed joint counts, and electronic messages to health care professionals (HCP). This work aimed to 1. Assess opinions of people with RA and health care professionals (HCPs) regarding design and functionality of an app and acceptability and usefulness of an app assisted telehealth approach for RA, 2. Develop an app, and 3. Assess usability of the app.

Methods

Interviews

Semi-structured interviews were undertaken with people with RA (American College of Rheumatology 2010 criteria) and HCPs, recruited from a hospital rheumatology service. Interviews explored technology use, app functionality, barriers and facilitators to app use, and potential impacts of app implementation on service provision and experience. Thematic analysis was performed and recruitment concluded when saturation was achieved.

App Development

An app (RAConnect) was developed in iOS and Android in co-operation with professional software developers. App design and function was informed by interview data, along with design principles informing high user-experience.

App Usability

People with RA recruited from the hospital clinic trialed RAConnect on their mobile devices for one month. App

usability was assessed using a “think aloud” interview post app download, an online survey which included the System Usability Scale (SUS) and free text feedback.

Results

Interviews

Nine people with RA (27-79yrs, 7 females, 1-26yrs of RA, low-moderate disease activity) and 11 HCPs were interviewed. Four themes were identified:

1. Variable app readiness.
2. Reduced barriers with high app usability and text message communication.
3. Pros and cons of PROs, with some ambivalence
4. Resource allocation and engagement - PROs reporting via an App is acceptable to guide HCPs in allocation of limited resource while also increasing patient engagement.

App Usability

Usability testing with 16 people with RA confirmed RAConnect had high usability (SUS 80/100, 90th centile). Both the app and a telehealth approach had high acceptability. “I think a wraparound thing like RAConnect would allow better spacing between appointments and for me it would make them less stressful.” 48 year-old Male. “I think that it has the potential to be very useful in managing my RA. It is easy to use and may mean less visits to the rheumatologist.” 49 year-old Male. “It’ll be a great tool for both the Doctors, their clinical staff and their patients.” 70 year-old Female.

Conclusions

A patient-held app for RA monitoring and communication with rheumatology care provider will be acceptable and desirable for many patients. Further research is required to develop multimedia material to train people with RA to perform joint counts and to validate patient-performed joint counts. During implementation assessment of the patient experience of using this telehealth system, disease outcomes, costs, and impact on workflow of HCPs will be required.

Address for correspondence

Rebecca Grainger: rebecca.grainger@otago.ac.nz

Competencies for Effective Use of Online Depression Information Among College Students

Yulong Gu^a, Zornitsa Kalibatseva^b, Xu Song^c

^aSchool of Health Sciences, ^bSchool of Social and Behavioral Sciences, ^cSchool of Arts and Humanities, Stockton University, Galloway, New Jersey, USA

Abstract

Mental health issues such as depression are rising among young people, who may benefit from online depression related information. However, the competencies required to search for and evaluate health information on the Internet were reported as poor among this population. We designed a web-based questionnaire survey including four case scenarios to understand how college students use online depression information, and what competencies or factors may facilitate effective use of the online resources.

Keywords:

Depression; Internet; information literacy.

Introduction

Depressive disorders are common among college students in the United States (US), with 14.9% students reporting a diagnosis of depression in their lifetime [1]. Depression is a disabling condition, and is sometimes accompanied by self-harming behaviors. Among college-age Americans, suicide has remained a leading cause of death in recent years (ranked No. 2 since 2013) [2]. Evidence-based treatments for depression exist, but are not accessed by majority of the college students with symptoms. Depression information services and programs on the Internet have been developed partly in response to this challenge. These online resources hold promise to reach college students, to support them in self-help, and to encourage them to seek professional help. But little is known about the effectiveness of online resources outside of the laboratory setting. Although young people regularly use and are generally satisfied with online mental health resources, more research is needed to determine if these resources effectively facilitate help-seeking behavior [3]. Moreover, college students may feel comfortable searching health information online, but their information literacy, in terms of searching for information and evaluating the information quality may be poor [4]. To understand how well college students can search the Internet for depression information and what relevant competencies they have, we designed a web-based questionnaire survey.

Methods

The questionnaire includes four realistic case scenarios to assess a participant's performance in finding and using online depression resources. In each case, the participant is asked to use the Internet to find answers to depression-related questions. We also ask participants to rate (in Likert scale) their health and information literacies. This questionnaire was piloted with two faculty and two students who confirmed its face validity.

In November 2016, 168 students enrolled in six undergraduate classes at a US university were invited via class group emails

to complete the web-based questionnaire survey. Additionally, 50 students were recruited through the university Psychology Research Pool. Data analysis aims to understand how well participants can use the Internet to answer questions, and what competencies / factors may facilitate effective use of online resources. The study protocol was approved by Stockton University Institutional Review Board.

Preliminary Results and Discussion

Among the initial survey respondents who completed the survey in the first week of the study period (N=149, median age=20; 79% female), 49% performed well in answering depression related questions in the case scenarios ($\geq 90\%$ accuracy), and 29% had good depression knowledge ($\geq 90\%$ accuracy). Less than half (44%) reported their health literacy level as moderate or higher (by familiarity with medical terms). More participants reported moderate or higher level of critical health literacy (73%) and information literacies (86%). The linear multiple regression model on all of the above knowledge and literacy factors indicates that the participants' performance in cases is associated with their depression knowledge ($p < .01$) and self-reported information literacy level ($p < .01$).

Mental health issues such as depression are rising, with depression being the second contributor to global burden of disease. Information on the Internet may support people dealing with depression. Preliminary results detected suboptimal health literacy and depression knowledge levels, and the students' depression knowledge and information literacy are significant predictors for effective use of the Internet. Thus, systematic training to develop these competencies may be necessary.

References

- [1] The American College Health Association, American College Health Association-National College Health Assessment Spring 2008 Reference Group Data Report (Abridged): The American College Health Association, *J Am Coll Health* 57 (2009), 477-488.
- [2] Centers for Disease Control and Prevention (CDC), Leading Causes of Death, in, CDC, Atlanta, GA, 2016.
- [3] S.D. Kauer, C. Mangan, and L. Sanci, Do Online Mental Health Services Improve Help-Seeking for Young People? A Systematic Review, *J Med Internet Res* 16 (2014), e66.
- [4] M. Stelfox, B. Hanik, B. Chaney, D. Chaney, B. Tenant, and E.A. Chavarria, eHealth Literacy Among College Students: A Systematic Review with Implications for eHealth Education, *J Med Internet Res* 13 (2011), e102.

Address for correspondence

Yulong Gu (Email: yulong.gu@stockton.edu)

Flexible Detection of Fall Events Using Bidirectional EMG Sensor

Hao Han^a, Xiaojun Ma^b, Keizo Oyama^c

^a The University of Tokyo Hospital, Tokyo, Japan,

^b The University of Tokyo, Tokyo, Japan,

^c National Institute of Informatics, Tokyo, Japan

Abstract

Falling is one of the most serious life-threatening events for the elders, and the growing population of elderly people motivates the development of ICT-based healthcare-oriented solutions for fall detection prevalently. In this poster, a bidirectional EMG (electromyographic) sensor network model is proposed for a more efficient and flexible detection of fall events based on simple communication between users and nursing care staff.

Keywords:

Fall Detection, Healthcare, Electromyographic Sensor

Introduction

As a nation running ahead of the curve in coping with the problems of an aging population, Japan has to face up the problem of a serious shortage of nursing care staffs in small and rural hospitals and healthcare facilities for the elderly. In this poster, we propose a wearable bidirectional EMG sensor network for flexible communication between users and nursing care staffs, and explain that this solution performs more realistically and rationally.

Wearable Bidirectional Sensor Network

For fall prediction, the analysis of health and medication data extracted from EHR (Electronic Health Record) is conducted [1]. However, it cannot realize a real-time prediction. Single-direction communication, such as surveillance camera and nurse call, is inefficient, and a simple "call/reply" interactivity is more effective in falls prevention [2] and could reduce workload of nursing staffs. We aim to detect falls with high positive predictive value and sensitivity, which means most real falls are correctly detected and less false falls are wrongly recognized. An interactive communication is provided between nursing care staffs, users, and the sensor network for releasing nursing workloads. When the fall prediction works, the users are soon warned, and when the fall detection fails, nursing staffs are called for helps through a light and easy-to-use interface.

An armband-type EMG sensor is imported to evaluate and record the electrical activity generated by skeletal muscles since arms give a sudden movement naturally and instinctively to stop falls. The sensor is also equipped with Wireless LAN, gravitational acceleration and gyroscope sensor. Through wearable sensor, the live logs are collected and stored in real-time. Live logs contain the continuous information of location (estimated by trilateration), status of movement and arm electromyogram data of users. The nursing staffs can learn the status of target users through the monitoring anytime. The collected live log is analyzed in real-time, and the fall alarm (location and user profile) is sent to nursing staffs if a fall event is detected (one of fall patterns is matched). Moreover, a vibration "call" (trigger built-in vibration motor) is remotely

sent to the sensor as a confirmation for a possible "reply" from the detected target user. The nursing staffs run to the fall location to help fallen users, or stop the help if they receive a cancellation "reply" from the user. Here, by repeating a predefined special hand gesture (e.g., hold fist and rotate) which is seldom occurred in general life and easily learnt by the elderly, users can send a "call/reply" message to the system as a simple and quick communication between users and nursing staffs when a wrong or missed detection happens. Moreover, based on the analysis of personal activity patterns, fall history records, profile information and various statistics (e.g., dangerous areas like stairs), a user-oriented customizable prediction mechanism is constructed. If a possible coming fall is predicted, a notification "call" is sent to the sensor as a warning message.

Results

Thirteen (13) panel users assisted our experimental data collection by wearing the armband sensor on their right forearms in activities of walking/sitting/falling/leaning (arms against wall to stop a fall). Through the raw data, we can find that the acceleration and gyroscope sensor data are not adequate enough to clearly recognize the falling since there were many slow-falling instead of sudden and fast falling. Therefore, the analysis of EMG data would contribute to a more precise fall detection as an important detection factor since there were violent fluctuations in the signals of sensors at muscle near ulna and radius when the users fell or almost fell (leaning). Here, positive predictive value = 81.8% and sensitivity = 69.2%, which performs better than sensors without EMG data.

Conclusion

A model of bidirectional EMG sensor network is designed to detect/prevent elder-oriented fall events, and EMG data was proved to be contributing factors of a more precise detection and prediction result. As future research, we plan to explore further the problems of fall pattern learning and recognition based on gathering more real data.

References

- [1] S. Yokota and K. Ohe, "Construction and evaluation of FiND, a fall risk prediction model of inpatients from nursing data", Japan Journal of Nursing Science, 2016.
- [2] R. A. Kenny, C. N. Scanaill, and M. McGrath, "Falls Prevention in the Home: Challenges for New Technologies", J. Soar, R. Swindell, and P. Tsang, Eds. Information Science Publishing, 2011.

Address for correspondence

Hao Han can be contacted at: han@hcc.h.u-tokyo.ac.jp

Design of Rehabilitation Treatment Coach Robot

Seung-Ho Han, Han-Gyu Kim, Chan-Yong Park, Ho-Jin Choi

School of Computing, KAIST, Daejeon, Korea

Abstract

In the rehabilitation treatment for at-home patients, it is not only costly for a patient to hire a professional therapist or personal trainer, but also time consuming for a therapist or trainer to visit all the patients. To improve the situation, we propose the rehabilitation treatment coach robot which helps the patient to do rehabilitation exercises alone. Designed economically with cheap parts, our robot provides multiple functions: rehabilitation program suggestion, rehabilitation posture correction, and emergency detection. A brief plan is presented for data collection and performance evaluation.

Keywords:

Exercise Therapy, Robotics, Neural Networks (Computer)

Introduction

Rehabilitation treatment helps patients with various physical or neurological diseases to restore physical or mental capabilities. It is important for the patient to practice consistently the given treatment until it is cured. In case of at-home patients, however, it is costly to hire professional therapists or personal trainers. Moreover, doing exercise alone can exacerbate the patient's condition if an improper posture or program is adopted. In order to help this situation, we propose a rehabilitation treatment coach robot. The robot not only helps the at-home patient to exercise alone, but also is designed to be cheap enough.

Methods

In order to design the robot as cheaply as possible, we use the most basic hardware, such as Microsoft Kinect, a display with touch sensors, a processor, a battery, and wheels with electric motors. Such rehabilitation coach robot has three functions: rehabilitation exercise suggestion, rehabilitation posture correction, and emergency detection.

The rehabilitation program suggestion module will recommend proper exercises to the patients for quick rehabilitation. The suggestions of rehabilitation program is mainly decided by communications with doctors.

In case of the posture correction module, the robot should be able to recognize the posture of the patient first. To do this, we use the Microsoft Kinect and its SDK that facilitates an extraction of the human skeleton data using built-in sensors. And then, since deep neural network (DNN) based posture recognition [1] is already shown good performance, we will train the model using deep learning algorithm to determine whether the posture is correct or not by using the skeleton data as an input to the model. Lastly, we will correct the posture of the patient using the output of the model.

In order for the robot to detect the emergency situation, we use the voice from the microphones and video input from cameras as an input to the DNN model [2]. Using this model, our coach

robot automatically recognizes the emergency situation and calls the doctor.

Experiment Plan

We plan to collect datasets for each module in the following ways. For the speech recognizer, the speech data is collected from at least 50 persons, and for posture correction module, the correct and incorrect postures are collected from professional trainers and ordinary persons. For emergency detection module, some emergency scenarios are designed and the datasets are collected from those scenarios.

We plan to use two evaluation criteria: automatic evaluation and human evaluation. For automatic evaluation, the criteria of evaluation is predefined for each module individually. For human evaluation, overall performance of the robot is evaluated by a group of therapists, trainers and patients.

Conclusions

In this paper, we have proposed a coach robot that helps and assists at-home patients during rehabilitation exercise alone. Designed with cheap parts, the robot has several main functions: exercise suggestion, posture correction, and emergency detection. We have presented our experiment plan for data collection and evaluation.

Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research, funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

References

- [1] A. Tang, K. Lu, Y. Wang, J. Huang, and H. Li, A real-time hand posture recognition system using deep neural networks, *ACM TIST* 6 (2015), 21.
- [2] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, Fall detection from human shape and motion history using video surveillance, in: *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, IEEE, 2007, pp. 875-880.

Address for correspondence

Seung Ho Han, School of Computing, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea, +82 01 8030 6802; Email: seunghohan@kaist.ac.kr

A Patient Centric Nursing Mobile App Redesign

Navas Hernán^a, Dozo Mariano^b, Dourado Ezequiel^a, Canteros Fernando^a, Rung Ingrid^a

^a Sanatorio Finochietto, Buenos Aires, Argentina,

^b Gizmedic, Buenos Aires Argentina.

Abstract

Sanatorio Finochietto has a nursing mobile app that allows safely preparing and administering medications. Hospitalized patients care experience is influenced by the system, the care process and nurse response, in this context, technology acts as a facilitator for care and humanized treatment. A "Patient-Centered Management System" would assist caregivers and patients to make individualized decisions about optimal care. This paper describe a paradigm shift to a mobile patient centered nursing application.

Keywords:

Electronic Health Record, Mobile Applications, Patient-Centered Care.

Introduction

Sanatorio Finochietto has a mobile solution for nurses that allow safe and fast administration of medicines and mobilize nursing processes [1]. A "Patient-Centered Management System" would assist caregivers and patients to make individualized decisions regarding optimal care for common clinical situations. It should also explicitly incorporate patient preferences, and would reinforce such decisions through patient focused performance measures.

This paper aims to describe a paradigm shift on a mobile app for nurses, initially developed to simplify nursing processes and redesigned to serve a patient centered model.

Methods

A multidisciplinary team was gathered to addressed this issue, including nursing leaders, user experience designer and developers experts, IT department members, and process specialists. The team analyzed nursing processes and tasks in different services of the institution, including interviews with users to know their opinions on the existing mobile app and their requirements, and observing users on their daily routines. The team also analyzed usage data from the existing mobile application (for drug administration) to identify heavy users and value their comments on the app [2].

After three years utilizing and improving the initial mobile app, the team focused on shifting from the nursing processes centric app to a patient centric version. To be able to achieve this objective, the patient will be the starting point of all nursing activities. Instead of being able to access all potential tasks, nurses will have a redesigned app which will display the list of assigned patients once the user logs in.

The beginning of the care process will be given by a portal that will provide information for each patient. This change will allow the nurse to have more information about the patient

before beginning the care process. With this, the personal context of each patient will be present in every interaction.

Results

The multidisciplinary team decided gradual deployment of the app. This allowed all stakeholders to participate on the process and be prepared for the changes the app brought. The gradual deployment started with functionalities related to information visualization. Later on simple records of information were incorporated, until completing all the services developed until today. The user's feedback and constant monitoring of the process led to issues related to daily use of this sort of app and to identify users' needs. The earned knowledge are the base of all the changes planned for the app redesign.

Conclusion

The formation of the multidisciplinary work team together with user's involvement throughout the process generated a positive feedback with the nursing area. This led to motivating the nursing team to participate actively on the deployment of new phases of the solution, facilitating the implementations.

The nursing mobile app new format should simplify how to adapt tasks and processes to different type of patients (adult in general admission, obstetric patient, adult intensive care unit, or neonate intensive care unit), guiding the user in the specific context of each patient and allowing for a more assisted process for the nurse while doing their tasks.

References

- [1] H. Navas, L. Graffi Moltrasio, F. Ares, G. Strumia, E. Dourado, and M. Alvarez, Using mobile devices to improve the safety of medication administration processes, *Stud Health Technol Inform* **216** (2015), 903.
- [2] J.E. Stevenson, G.C. Nilsson, G.I. Petersson, and P.E. Johansson, Nurses' experience of using electronic patient records in everyday practice in acute/inpatient ward settings: A literature review, *Health Informatics J* **16** (2010), 63-72.

Address for correspondence

Dr. Hernán Navas (hnavas@sanatoriofinochietto.com)

Connecting PHRs and EHRs for a Sustainable National Health System

Evelyn Hovenga RN PhD FACHI FACS FACN, Heather Grain A Dip HIM G Dip IS MHI FACHI

eHealth Education Pty Ltd, East Melbourne, Victoria, Australia

Abstract

An EHR for integrated care (IEHR) is defined by the International Organization for Standardization (ISO) [1]: "...a repository of information regarding the health status of a subject of care, in computer processable form, stored and transmitted securely, and accessible by multiple authorised users, having a standardized or commonly agreed logical information model that is independent of EHR systems and whose primary purpose is the support of continuing, efficient and quality integrated health care. It contains information which is retrospective, concurrent and prospective." We need to differentiate between EMR/EHR and the lifelong PHR in terms of type of data storage, sharing and use [2-3].

Keywords: Medical Records, Integrated Health Care Systems, Personal Health Records

- National policy evaluation and development
- Personal health observations incl. fitness, wellness monitoring.

• PHR	←→	EMR/EHR
Individual controlled health record	Information exchange or shared use, under mixed governance models	Healthcare provider controlled health record

Figure 1: PHR-EMR continuum

There need to be stringent data management requirements for all data managed by healthcare providers to ensure data integrity, completeness, meet medico-legal requirements and to protect confidentiality. Connectivity and interoperability continue to be huge challenges to be addressed globally.

Introduction

EMR/EHR: episodic medical/health record documenting all events managed by one healthcare service provider in real time and in sufficient detail enabling decision making.

- Data sharing and secondary data use purposes:
 - With individual's PHR enabling continuity of care across providers and locations over a lifetime, including episodic care and support service co-ordination.
 - Clinical trial research
 - Organisational/individual provider performance monitoring
 - Quality and evidence based practice evaluation by case type and service/treatment delivery methods between organisations.
 - Organisational resource management
 - Monitoring service demand and manage in-house service scheduling
 - Regional and national reporting to support funding, workforce planning, policy evaluation and development based on demographics services and disease burden.
 - Link with supply chain, prosthesis, medications, equipment and supplies used.

PHR—personal lifelong record used to document and record key events and outcomes.

- Data sharing and secondary data use purposes:
 - Personal management of required collaborative and/or ongoing services, incl scheduling appointments, identifying availability of care/service options by location.
 - Sharing information with new treatment/care/service providers
 - Public and population health management and global monitoring

Methods, Results, and Conclusions

Connectivity and Semantic Interoperability

- Agreed platform & national reference architecture [4]
- Agreed set of datatypes
- Well managed unique identifiers
- A standard 'language' & terminologies
- Data sharing agreements between custodians
- Governed agreed minimum data sets.
- Governance of clinical knowledge stored in computable formats.
- Global equipment/supplies catalogue standard
- Legislation and regulation supporting these interoperability principles.

References

- [1] ISO/TR 20514:2005 *Health Informatics- EHR Definition, Scope and Context*.
- [2] E.J S Hovenga, S.Heard eHealth Records and Future Healthcare in E.Hovenga, M.Kidd, S.Garde, C.Hullin *Health Informatics: an overview*, IOS Press, Amsterdam.2010
- [3] E.Hovenga, H. Grain An Information Paradigm Shift is Required to realise EHR Benefits, in *Medinfo2015 Proceedings*, IOS Press Amsterdam 2015
- [4] openEHR wiki <https://en.wikipedia.org/wiki/OpenEHR>

Address for correspondence

Corresponding author Dr Evelyn J.S Hovenga, Melbourne, Australia.
ehovenga@gmail.com

Intelligent Digital Environment for Wellbeing and Health

Inga Hunter^a, Caroline Lockhart^a, Phoebe Elers^a, Dick Whiddett^a, Hans Guesgen^b

^a School of Management, Manawatu Campus, Massey University, Palmerston North, New Zealand

^b School of Engineering and Advanced Technology, Manawatu Campus, Massey University, Palmerston North, New Zealand

Abstract

This three-phase project seeks to identify the user-requirements and develop a novel and unobtrusive technology platform interpreting and integrating diverse sources of information from a variety of digital devices monitoring the health and wellbeing of an older person with social media networking. This technology platform will communicate that information within an individualised support network thus supporting the older person to 'age in place'. This poster presents the process and preliminary results of phase one.

Keywords:

Technology; Delivery of Health Care; Social Networking.

Introduction

Developed countries are experiencing an ageing population, often with complex health problems, placing unsustainable demand on health care services [1]. Staying at home in age-friendly communities supported by primary health care and family (age-in-place) is a preferred option of older people, can provide better quality of life and be a more sustainable model of care enhanced by information technology [1, 2].

Smart home information technologies (sensors, recording devices, etc.) create intelligent environments that monitor and control the environment, track activities of daily living, and recognise significant departures from normal behaviour [3]. Social media networks offer many-to-many (M:N) relationships between the older person, family/friends, aged care workers and health providers, widening and individualising the care/support network of the older person.

Despite the relative maturity of smart home and social media technologies, they have not been integrated within the 'ageing in place' model of care and there is a lack of research into the socio-technical aspects of information sharing using social media that this project seeks to fill [4, 5].

Methods

A literature review identified key questions for an expert group workshop. The workshop was conducted with 24 attendees of the annual conference of Health Informatics New Zealand, a member society of IMIA. A many:many (M:N) collaborative information flow model was developed to facilitate discussion (Figure 1).



Figure 1 - A many:many (M:N) collaborative information flow model

The outer ring represents health providers, the middle ring the older person's support network and the inner ring the older person. Information can flow within and between each ring in any direction. Responses were analysed by thematic analysis.

Results

Key themes for user requirements were identified:

1. Data and Information – methods of data retrieval, processing and interpretation;
2. Connection and Collaboration – impact of digital connection on education, wellness and user roles;
3. Autonomy and Control – technologies' impact on user autonomy and control.

The respondents identified that sharing information on health and wellbeing to identify early deterioration in health status was more important than sharing specific illness information.

Conclusions

Existing literature primarily focuses on the problems with smart home/social media technology adoption. The workshop themes suggest user requirements to resolve these problems which will be further investigated in phase two with semi-structured interviews/focus groups with older people, members of their support network and health providers. Phase three will use a co-design with end-user development process.

Acknowledgements

Funded by Health Research Council, New Zealand, ref 16/679

References

- [1] Ministry of Health. *Health of Older People Strategy: Consultation draft*. Wellington: Ministry of Health, 2016.
- [2] Labour, the Greens and Grey Power. *A Report into Aged Care*, New Zealand House of Representatives, 2010.
- [3] F. Rivera-Illingworth, V. Callaghan, and H. Hagrais, Detection of Normal and Novel Behaviours in Ubiquitous Domestic Environments. *The Computer Journal* 53(2) (2007).
- [4] M. Chan, D. Esteve, C. Escriba, and E. Campo, A review of smart homes - present state and future challenges. *Computer methods and programs in biomedicine* 91 (2008), 55-81.
- [5] B. Hughes, I. Joshi, and J. Wareham, Health 2.0 and Medicine 2.0: Tensions and Controversies in the Field. *Journal of Medical Internet Research* 10(3): e23 (2008).

Address for correspondence

Dr Inga Hunter email: i.hunter@massey.ac.nz

The Development of mWellcare, an mHealth System for the Integrated Management of Hypertension and Diabetes in Primary Care

Devraj Jindal ^a, Priti Gupta ^a, Dilip Jha ^a, Vamadevan S Ajay ^a, Pramod Jacob ^a, Kriti Mehrotra ^b, Pablo Perel ^c, Jonathan Nyong ^c, Ambuj Roy ^d, Nikhil Tandon ^d, Dorairaj Prabhakaran ^a, Vikram Patel ^c.

^a Public Health Foundation of India, New Delhi, India

^b Dimagi Software Innovation Private Limited, New Delhi, India

^c London School of Hygiene and Tropical Medicine, London, UK

^d All India Institute of Medical Sciences, New Delhi, India

Abstract

This paper describes development of mhealth application “mWellcare” for the integrated management of hypertension and diabetes in public primary health care settings. mWellcare application was developed in four phases: identifying gaps in usual care; identifying components of intervention; developing intervention; evaluating acceptability and feasibility through pilot testing. Final version of mWellcare application is capable of computing personalized evidence-based management plan for hypertension, diabetes and co-morbid conditions (depression and alcohol use disorder).

Keywords:

Decision Support Systems, Clinical; Primary Health Care; Chronic Disease.

Introduction

Cardiovascular diseases (CVD) and diabetes are among the leading causes of premature adult deaths in India [1]. Systematic reviews has shown that mHealth based Decision Support System (DSS) improves preventive care and physician’s clinical decision quality in hypertension and diabetes management[2]. The aim of this paper is to describe the steps and processes in the development of mWellcare, a complex intervention based on mHealth technology.

Methods

We used the framework for the development of ‘complex’ interventions proposed by the Medical Research Council (MRC), United Kingdom[3]. Based on this framework, we followed these steps: 1) Identifying the gaps in usual care; 2) Identifying the components of intervention; 3) Development of intervention components; and 4) Evaluating acceptability and feasibility through pilot testing.

Results

The mWellcare system platform is built with Python using django web framework, mobile application is driven by an XML application configuration layer with JavaRosa at the core and database is Couch DB. Standards used for its development are ICD 10 for diagnosis, LOINC for lab results,

UIDAI (Unique Identity Authority of India) Aadhaar for patient identification, HL7 for messaging, CCD for clinical document exchange. It is used by primary care nurses and doctors and has the ability to: store and integrate longitudinal health records electronically; provide automated guideline-recommended treatment plan and life style advices tailored to patient clinical profiles; enables longitudinal patient monitoring and alerts to the need for changes in management; send out automatic short-messaging services (SMS)reminders and alerts to patients ; serves as data collection tool for remote quality assurance of NCD programs; and exports standards compliant healthcare data to relevant receiving healthcare systems. mWellcare system is currently in use in 40 community health centers in two states of India, Haryana in the north and Karnataka in the south.

Conclusions

We have described the process of development of a mHealth intervention seeking to improve the integrated management of hypertension and diabetes in primary care level. The effectiveness and cost-effectiveness of the intervention is currently being evaluated in a cluster randomized controlled trial in India (trial registration number NCT02480062).

Acknowledgements

This work was supported by the Wellcome Trust (Grant Number 096735/A/11/Z).

References

- [1] *Global Status Report on noncommunicable diseases*, in, World Health Organization, Geneva, 2014.
- [2] A.X. Garg, N.K. Adhikari, H. McDonald, M.P. Rosas-Arellano, P.J. Devereaux, J. Beyene, J. Sam, and R.B. Haynes, Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review, *JAMA* **293** (2005), 1223-1238.
- [3] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, and M. Petticrew, Developing and evaluating complex interventions: the new Medical Research Council guidance, *Int J Nurs Stud* **50**, 587-592.

Address for correspondence

Priti Gupta, email: priti@ccdcindia.org

How Are Doctors Using Mobile Electronic Medical Records? An In-Depth Analysis of the Usage Pattern

Juntae Kim^a, Yura Lee^b, Sanghee Lim^c, Jae-Ho Lee^{b,d}

^aKAIST 85 Hoegiro, Dongdaemun-gu, Seoul Korea,

^bDepartment of Biomedical Informatics, Asan Medical Center 88 Olympic-ro 43-gil, Songpa-gu, Seoul Korea

^cThe Johns Hopkins Carey Business School 100 International Drive, Baltimore, MD 21202, USA

^dDepartment of Emergency Medicine Asan Medical Center 88 Olympic-ro 43-gil, Songpa-gu, Seoul Korea

Abstract

In recent years, major hospitals have started using mobile electronic medical records (m-EMR). However, usage patterns based on actual log data have rarely been discussed. In this study, we investigated usage patterns of a m-EMR in several dimensions to understand the m-EMR in depth.

Keywords:

Telemedicine, Electronic health records, Hospitals, Clinical laboratory information systems

Introduction

Recently, major hospitals have started using mobile electronic medical records (m-EMR) in practice; evidence suggests that they are underutilized [1]. Prior studies primarily focused on the design of m-EMRs and the behavioral factors that influence their adoption. Few studies have investigated use patterns using an actual m-EMR record. An in-depth understanding of these patterns based on real logs is important for illustrating the value of using an m-EMR. We aimed to scrutinize the usage patterns of m-EMR by investigating usage patterns of an m-EMR designed to access hospital information systems (e.g., Picture archiving and communication system and laboratory information system).

Methods

We examined an m-EMR with 13 menus used by the biggest hospitals in South Korea. We collected 524,345 records from 653 doctors (272 residents, 174 fellows, and 246 professors) March 2015- December 2015. We excluded 625 logs for two non-physician menus. We performed descriptive analysis in five perspectives: hourly units, daily units, user positions (rank), departments, and menus.

Results

The frequency of use differed between time units and days (Figures 1-2); 2) The frequency of use depends on doctor rank and department type; and 3) usage distributions of each function by department and time unit were found.

Discussion

Doctors used the m-EMR early in the morning (6-9 am) and most often on Sundays. An m-EMR was used when the doctor could not access EMR through PC. The usage by resident surgeons was very low. Investigating factors related to the low use of m-EMR by resident surgeons is needed. Since menus can be interrelated, we expect that clustering analysis will provide useful insights.

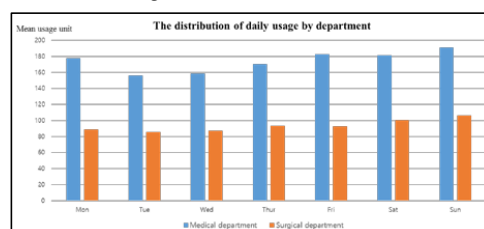


Figure 1. The distribution of daily usage by department

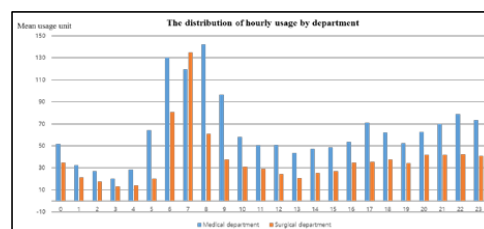


Figure 2. The distribution of hourly usage by department

Conclusion

The use of m-EMR in clinical practice was early and there was little evidence of successful assimilation of m-EMR in major hospitals. Thus we believe that much research is needed to improve the use and functions of m-EMR.

References:

- [1] B. Middleton, Achieving U.S. Health information technology adoption: the need for a third hand, *Health Aff (Millwood)* **24** (2005), 1269-1272.

Address for correspondence

Jae-Ho Lee, Department of Emergency Medicine Asan Medical Center 88 Olympic-ro 43-gil, Songpa-gu, Seoul Korea, rufiji@gmail.com

A Content Analysis of Mobile Apps for Chronic Kidney Disease Patient Care: Searching in English and Chinese

Ying-Li Lee^{a,b}, Yan-Yan Cui^b, Polun Chang^b

^a Department of Nursing, Chi Mei Medical Center, Tainan, Taiwan, China

^b Institute of Biomedical Informatics, National Yang Ming University, Taipei, Taiwan, China

Abstract

Chronic kidney disease (CKD) is a global public health issue. Mobile technology is pervasive and widely used in chronic disease care. More and more, CKD mobile applications (apps) can be found on popular mobile application platforms, especially in Chinese. We aim to explore current mobile apps for CKD patient care (searching in English and Chinese) through content analysis to identify the app functions that health professionals can use in CKD patient care and to help CKD patients with self-management.

Keywords:

Mobile Applications; Renal Insufficiency, Chronic; Self Care

Introduction

The prevalence of CKD worldwide is on average around 13.4% (11.7-15.1%) [1]. According to Diamantidis and Becker, patients with CKD would ideally have confidential and convenient mobile access to their therapy plans and medical records [2]. Various CKD-related mobile apps in different languages can be found on several popular mobile application platforms. We aimed to explore currently available mobile apps for CKD patient care (searching in English and Chinese) through content analysis, a common method for exploring the content and functions of mobile apps, to identify functions that health professionals can use in CKD patient care and to help CKD patients with self-management.

Methods

Two researchers with prior training in coding principles searched the Apple store, Google Play, and 360 mobile assistant using the keywords "Kindey", "Renal", "Nephro", "Chronic Kidney Disease", and "CKD". We included apps that were related to CKD care and that could be used on smartphones or tablets, but excluded journals, medical conference manuals, hemodialysis-related or inoperable apps. The content and functions of the programs were coded separately by two people. The target users (health professionals, patients or both) were determined according to the actual function content. Two coders compared and discussed the coding results to achieve consensus. If no consensus was reached, a third researcher would participate in the discussion. After coding, descriptive statistics were used to summarize the content and functions of the apps. All analyses were performed using SPSS 20.0 statistical software.

Results

As of June 5, 2016, we identified 105 apps available from the Apple store, 130 apps on Google play, and 69 apps on the 360 mobile assistant. After deduplicating, 189 apps were remaining.

Following exclusion based on default criteria, 166 CKD-related apps were left for the analysis. In all, 27.11% of the reviewed apps had a final release in a recent year. The main app language was English (51.81%), followed by Simplified Chinese (18.67%), and 11 apps that were multilingual. Also, 69.28% apps were free, 19% had privacy claims, and 39.76% required registration.

Based on the functionality and content of the 166 apps we reviewed, there are 64 that target health professionals as users. The most common functions are clinical calculation (including eGFR) (39.06%), eGFR calculation (21.88%), CKD medical professional information (20.31%) and CKD care decision support (18.75%). Only 6 apps included patient management functions (9.38%), and two apps provided tele-consultation. There were 59 apps suitable for patients. The most common app functions include CKD nutrition care information (64.41%), self-management (35.59%), online consultation (27.12%) and CKD nutrition management (25.42%). There were 43 apps targeted at both health professionals and patients. The main functions were simple eGFR calculators (74.42%), CKD triage (27.91%) and CCR calculations (23.26%). There were five pairs of Simplified Chinese language APPs developing into patients' and physicians' APPs to provide interaction between patients and physicians. None of the apps qualified as a medical device according to MEDDEV 2.1 / 6 Guidelines.

Conclusions

The study shows that the functions of current CKD mobile apps are to support various types of numerical calculation and provide information about disease and nutrition care. The Simplified Chinese language APPs were developed to support patient and physician interaction. We suggest that CKD mobile apps be designed to protect the confidentiality of data, to assist patients in self-management, and to aid health professionals with individual patient management.

References

- [1] N.R. Hill, S.T. Fatoba, J.L. Oke, J.A. Hirst, C.A. O'Callaghan, D.S. Lasserson, and F.D. Hobbs, Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis, *PLoS One* **11** (2016), e0158765.
- [2] C.J. Diamantidis and S. Becker, Health information technology (IT) to improve the care of patients with chronic kidney disease (CKD), *BMC Nephrol* **15** (2014), 7.

Address for correspondence

Polun Chang, Institute of Biomedical Informatics, National Yang Ming University, No.155, Sec.2, Linong Street, Taipei, 112 Taiwan; polun@ym.edu.tw

Developing Mobile Health Management System for Patients with Musculoskeletal Tumor

Nai-Wen Liang^a, Hung-Yi Liu^a, Chun-En Chien^a, Qi-Guang Chen^a,
Po-Lun Chang^a, Wei-Ming Chen^b, Bo-Gui Wu^b

^a Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

^b Division of Joint Reconstruction, Department of Orthopedics, Taipei Veterans General Hospital, Taipei, Taiwan, China

Abstract

This reaserach provided a postoperative system for osteosarcoma which is seperated into two parts, Bone Patient and Bone Care. We focus on using smartphone features and proper interface design to facilitate self-monitoring, health analysis, and to make appointments. It achieves the effects of shared decision making (SDM). During the two weeks of research, we achieved satisfied with 4.03 ± 0.67 TAM score..

Keywords:

Telemedicine; Osteosarcoma; Decision Making.

Introduction

Osteosarcoma patients are more likely to be cured if they are cancer-free five years after diagnosis or if they have completed chemotherapy, including regular physical examinations and/or medical tests [1]. In the past, inefficient phone interviews were used to track patients' situations.

Methods

In this study, we developed two functions: Bone Patient, for osteosarcoma patients and Bone Care, for case managers (Figure 1).

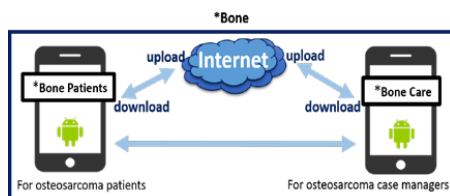


Figure 1– The structure of system(Bone)

We used Android “navigation drawer” and “card collection”. As a server recieves the record, it is saved into a MySQL database. Bone Patient includes file uploading, follow-up visit, inspections, self-evaluation, questionnaires, health education, and related links. Bone Care includes to-do-lists and completed tasks. We applied primary data collection with the use of judgment sampling, like basic personal information about patients, the test of TAM and integrated advice.

Results

The system made it easier for patients to make appointments. Also, the camera and voice recorder helped to describe symptoms and pain. It made it simpler for case managers to arrange and sort appointments by date or urgency (Figure 2).



Figure 2 – From left to right, the follow-up visit page and the to-do-list page

In Bone Patient, the highest score was for the learnability of the system (4.38 ± 0.61); the lowest scores for reliability of the system (3.52 ± 0.78). The total score in Bone Patient was 4.03 ± 0.67 . Most of the testers' agreed or strongly agreed with the use of Bone Patient and Bone Care.

Conclusions

This study focused on the development of a post-surgery tracking system for patients with musculoskeletal tumors. It makes patient care conveniently accessible through a smartphone.

Acknowledgements

This research was funded by the Department of Orthopedics at Taipei Veterans General Hospital.

References

- [1] C. Gerrand, N. Athanasou, B. Brennan, R. Grimer, I. Judson, B. Morland, D. Peake, B. Seddon, J. Whelan, UK guidelines for the management of bone sarcomas, *Clin Sarcoma Res* 6 (2016):7.

Address for correspondence

Polun Chang PhD
Institute of Biomedical Informatics, National Yang-Ming University
Address: No. 155, Sec. 2, Linong St., Beitou District, Taipei City 112, Taiwan
E-mail: polunchang@gmail.com
Phone: 886-2-28267238

Setting the Agenda for Personal Health Records in Low- and Middle-Income Countries

Thomas Mwogi^{a,b} and Martin C. Were^{a,c,d}

^a Institute of Biomedical Informatics, Moi University, Eldoret, Kenya,

^b Moi Teaching & Referral Hospital, Eldoret, Kenya,

^c Department of Biomedical Informatics & Medicine, Vanderbilt University, Nashville, TN, USA,

^d Vanderbilt Institute of Global Health, Vanderbilt University, Nashville, TN, USA

Abstract

In the developed world, Personal Health Records (PHRs) have been demonstrated to improve patient adherence, reduce medical errors, improve patient-provider communication, improve chronic disease management, and promote behavior change. PHRs have not been widely adopted in low- and middle-income countries. There is rising use of smartphones, adoption of national-level electronic health systems, and change in historical perceptions on acceptance of technology. It is thus an opportune time to critically examine the potential role and approaches to PHRs in LMICs.

Keywords:

Health Records, Personal; Telemedicine; Smartphone

Introduction

In developed countries, PHRs have been demonstrated several benefits [1, 2]. The literature describes three predominant models of PHRs implementations, namely: Stand-alone, Web-based, and Integrated PHRs.

In general, Integrated PHRs show greater benefits when compared to the other types of PHRs [2]. In developed countries, PHRs are in wide use, but these systems are yet to be widely adopted in Low- and Middle-Income Countries. Several factors make this an opportune time to reconsider the role and feasibility of PHRs in LMICs. In particular, the exponentially increasing ownership of powerful smartphones and tablet devices, and the widespread and national-level adoption of Electronic Health Record systems (EHRs) make it highly feasible to adopt integrated PHRs that can be availed to individuals within LMICs.

An Approach to Implementing PHRs in LMICs

PHRs implementations in LMICs need to be grounded on rigorous evidence. This is particularly important given that implementation can introduce new challenges around data privacy, security, literacy, and patient-provider relationships [1].

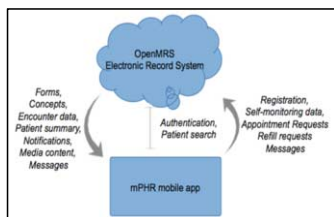


Figure 1 - Integration of mPHR mobile app with the OpenMRS

Given that integrated PHRs provide the most benefit, it is highly recommended that emerging PHRs for use within LMICs are integrated with nationally-endorsed EHRs. Figure 1 provides a schematic of such an integration with the OpenMRS EHRs that is being widely deployed in LMICs. Core to the

PHRs agenda is to define perceptions on PHRs and desired core features (Table 1). There also needs to be a multi-dimensional and comprehensive approach to the agenda for implementing PHRs (Table 2).

Table 1 – Sample of initial core PHRs features

Patient Reminders / Notifications (e.g. appointments)
Patient-Provider Messaging & Consultation function
Patient Clinical Summary
Laboratory Results Viewer
Health Educational Media
Facility Locator
Self-monitoring Tool (e.g. track blood sugars)

Table 2 – PHR Agenda for LMICs

Policy	Who owns the data, WHO Guidance
Technology Agenda	Usability, user centered, data size consideration, Operating systems, HIEs
Sustainability Agenda	Insurers, Mobile operators, Value added services, Public-Private Partnerships
Evidence	Benefit vs Cost evaluations,

Conclusion

A convergence of factors make this an opportune time to implement PHRs in LMICs. Clear policy, technology, sustainability and evaluation agenda are key to realizing the potential of PHRs in LMICs.

Acknowledgements

This work was supported in part by the NORHED program (Norad: Project QZA-0484). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Norwegian Agency for Development Cooperation. All authors report no conflicts of interest.

References

- [1] D.C. Classen, Clinical decision support systems to improve clinical practice and quality of care. *Journal of American Medical Association* 280(15) (1998), 1360-1361.
- [2] P.C. Tang and et al., Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *J Am Med Inform Assoc* 13(2) (2006), 121-126.

Address for correspondence

Thomas S. Mwogi, MMed, MS

thomasmwogi@mtrh.go.ke

Empathy Bot: Conversational Service for Psychiatric Counseling with Chat Assistant

Kyo-Joong Oh, DongKun Lee, ByungSoo Ko, Jonghwan Hyeon, Ho-Jin Choi

School of Computing, KAIST, Daejeon, Korea

Abstract

There are earlier studies for psychiatric counseling using chat bots. These studies have not considered the user's emotional status and ethical judgment to provide interventions. This paper proposes an intelligent assistant for psychiatric counseling that understands dialogues using high-level features of natural language understanding, and multi-modal emotion recognition. A response generation model using machine learning provides suitable responses for clinical psychiatric counseling.

Keywords:

Psychiatric Nursing; Emotional Intelligence; Natural Language Processing

Introduction

Chat bot is intelligent assistant software that responds to the users as conversational interaction. In an early study, Roberts et al. have developed a chat bot providing interventions to change subsequent behaviors for young adults who have alcohol risk [1]. Elmasri and Maeder provide emotional responses for mental health care according to user's behaviors and dialogues with a online chat bot [2].

In this paper, we propose a chat assistant called "Empathy Bot" which provide interventions and emotional responses for mental health care in a chat bot platform. We develop a deep interaction model that understands emotions and dialogues based on emotion recognition and natural language processing.

Methods

Empathy Bot consists of two parts: multi-modal emotion recognition, and natural language processing.

Multi-modal Emotion Recognition

To provide psychiatric counseling, we have to recognize user's emotional status in communication through text, voice, and visual expression. To recognize user's emotion, the service applies multi-modal emotion recognition methods from dialogues, intonation, and facial expression. The methodologies enable continuous observation of emotional changes sensitively.

Natural Language Processing

The empathy bot, which provides conversational service for psychiatric counseling collects and summarizes the dialogues of text. For the intelligent response we apply case-based reasoning, continuous long-term monitoring, and ethics judgment, etc. As to response generation for psychiatric

counseling, we apply a hybrid of template-based and machine learning-based approaches.

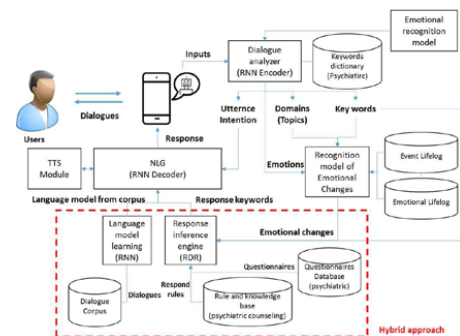


Figure 1– The Architecture of Empathy Bot

Conclusion

We have presented a chat bot called "Empathy Bot". More accurate emotion recognition based on multi-modal approach and appropriate responses based on clinical psychology with ethical judgment showed better satisfaction to users who need psychiatric counseling for mental health care.

Acknowledgements

This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean Government (MSIP) (No.2013-0-00131, Development of Knowledge Evolutionary WiseQA Platform Technology for Human Knowledge Augmented Services).

References

- [1] L. W. Roberts, J. Battaglia, and R. S. Epstein, Frontier eth- ics: Mental health care needs and ethical dilemmas in rural communities, Psychiatric Services, 1999.
- [2] D. Elmasri and A. Maeder, A Conversational Agent for an Online Mental Health Intervention, In: International Con- ference on Brain and Health Informatics, 2016, pp. 243-251.

Address for correspondence

Kyo-Joong Oh, School of Computing, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea. Phone: +82-10-9379-1386; Email: aomaru@kaist.ac.kr

Factors Influencing Progress of Health Information Exchange Organizations in the United States

Lauren M. Overhage^{a,e}, Jennifer Covich-Bordenick^b, Xiouchun Li^{c,e}, J Marc Overhage^{d,e}

^a Adler University, Chicago, IL, US, ^b eHealth Initiative, Washington, DC, USA

^c Indiana University School of Medicine, Indianapolis, IN, US, ^d Cerner Corporation, Kansas City, MO, USA

^e Regenstrief Institute, Indianapolis, IN, USA

Abstract

Progress is being made toward improved healthcare interoperability in the United States, but exchange between electronic health records alone is insufficient. Using data from the eHealth Initiative's Annual Survey of Health Information Exchange, we developed models of HIE financial and operational progress. Our analysis suggests that organizations that focus on enabling exchange through education and policy need to be considered separately from those focused on the actual exchange. The associations between characteristics and progress in data exchanging HIEs suggest that diversity of participants as both originators and receivers of data and breadth of data are important underlying success factors.

Keywords:

Health information exchange; Program evaluation; United States

Introduction

Health Information Exchange organizations (HIEs) have emerged as one component of a multifaceted approach to enabling interoperability in the US. Some, however, have questioned the viability of HIEs [1, 2]. In order to understand the factors that may be important in sustainability of HIEs, we undertook an analysis of data from the eHealth Initiative's annual survey of organizations and initiatives working in the area of health information exchange in order to validate other's findings and to model the progression of HIEs using longitudinal, contemporaneously collected data.

Methods

We used data from the eHealth Initiative's (eHI) annual survey on Health Information Exchange. We used responses to the questions which were the same or, that we judged to be sufficiently similar, from the surveys in 2006, 2007 and 2008 – a total of 334 in all. In order to establish useful and meaningful taxonomy of HIEs, we reviewed previous publications describing HIEs, created multidimensional visualizations and then used a modified Delphi process to define taxa.

We grouped initiatives reporting their progress stage as 1 through 3 as not yet operational; those reporting stage 4 as pilot and HIEs that characterized themselves as stage 5 through 7 as operational. Because HIEs might be optimistic in their self-classification, we adopted a data driven definition as well. We calculated Spearman correlations using SAS[®] to find correlation coefficients for the predicted stages of HIEs.

Results

Based on our definitions we found that only 117 (76%) of the 154 HIEs responding to the survey we focused on exchanging data and of those only 42 (27%) were operational in 2008. Progress is, however, not linear. Only a minority (16%) of HIEs regressed year to year with approximately 40% progressing. A large majority of the HIEs were able to correctly predict their own developmental stage on eHI's maturity model for the upcoming year with Spearman correlation coefficients of 0.89, 0.88 and 0.93 for 2006, 2007 and 2008 respectively. The variables with the highest odds ratios in association with financial progression from highest to lowest were hospital as a receiver of data, one time or recurring fees from participants, receiving grants, laboratory or radiology as a data provider and ambulatory physician as a receiver (all $p < 0.01$). The variables associated with operational progression were hospital as receiver, ambulatory physicians as receivers, laboratory or radiology as a data provider, one time or recurring fees, having received grants and sum of the unique types of participants (all $p < 0.01$).

In the multivariate analysis, using all the independent variables of interest to predict the two outcomes, none of the independent variables had significant impact. When we examined pairwise correlations among the independent variables most of them are significant, which indicates strong multi-collinearity (highly correlated predictor variables) making simple interpretation of the regression coefficients as measuring marginal effects from the multivariate model unwarranted.

References

- [1] J. Adler-Milstein, S.C. Lin and A.K. Jha, The Number of Health Information Exchange Efforts is Declining, Leaving the Viability of Broad Clinical Data Exchange Uncertain, *Health Affairs* **35** (2016), 1278-1285.
- [2] J. Adler-Milstein, D.W. Bates and A.K. Jha, Operational Health Information Exchanges Show Substantial Growth, But Long-Term Funding Remains a Concern, *Health Affairs* **32** (2013), 1486-1492.

Address for correspondence

J. Marc Overhage, MD, PhD, VP & Chief Medical Informatics Officer, Cerner Corporation, 51 Valley Stream Parkway, Malvern, PA, 19355 e-mail (preferred): marc.overhage@cerner.com

3D Documentation of Chronic Wounds Using Low-Cost Mobile Devices

Ekaterina Sirazitdinova^a and Thomas M. Deserno^b

^a Department of Medical Informatics, Uniklinik RWTH Aachen, Aachen, Germany

^b Peter L. Reichertz Institute for Medical Informatics, University of Braunschweig, Braunschweig, Germany

Abstract

The state-of-the-art wound assessment is manually performed by clinicians. This is ineffective and imprecise. Special devices have been developed for automation, but they are expensive and often lack portability. We present a concept of a low-cost system for wound documentation and analysis. Using computer vision techniques, 3D reconstructions of wounds can be created from images taken with a simple camera, such as those embedded in mobile phones.

Keywords:

Wound Healing, Documentation, Imaging, Three-Dimensional

Introduction

We aim to provide a new standard for documentation of chronic wounds by introducing an automatic 3D method for wound reconstruction followed by automatic analysis of the lesion. Our goal is to make the technology inexpensive, connective, and available for broader use.

Method

Images and inertial measurement unit (IMU) data are recorded using the interactive mobile app (Figure 1).

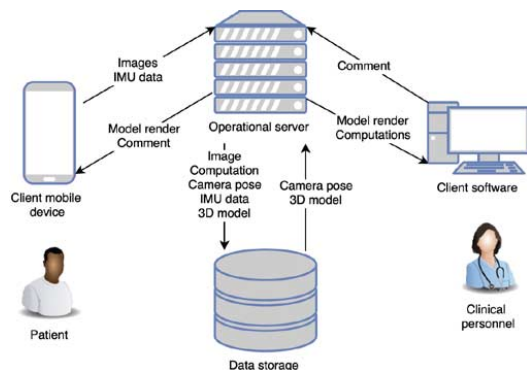


Figure 1 – System architecture

For 3D reconstruction, we adapt a method, where dense tracking is performed in each recorded frame [1]: camera positions are estimated using a dense feature-free monocular camera tracking and corrected using the information from the IMU. Based on the overlap between images, selected keyframes are rectified according to the camera position of incoming frame and dense stereo matching is performed using stereo block matching. The acquired depth maps are integrated into a volumetric model using the KinectFusion algorithm [2].

Result

Our system design is composed of two loops (Figure 2).

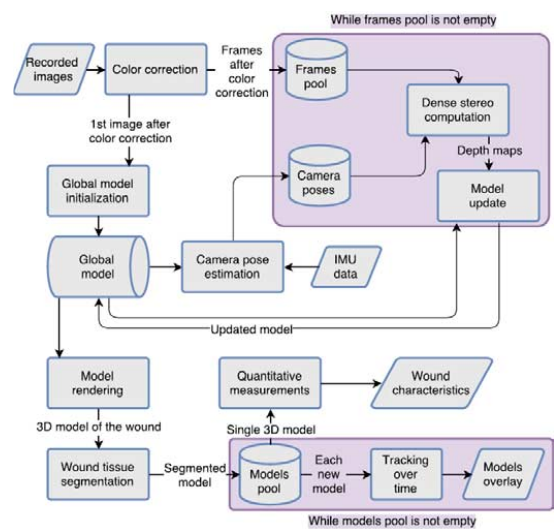


Figure 2 – Processing chain

Conclusions

A concept of a system using inexpensive consumer-level hardware for 3D wound reconstruction is presented. A novel approach to 3D reconstruction using mobile devices equipped with a single RGB camera is adopted into the medical domain. Accompanied with color correction, tissue segmentation, and tracking over time, the proposed system design is capable of providing objective and quantitative measurements of wounds.

References

- [1] P. Ondruska, P. Kohli, and S. Izadi, MobileFusion: real-time volumetric surface reconstruction and dense tracking on mobile phones. *Proc ISMAR 2015* (2015).
- [2] R.A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A.J. Davison, P. Kohi, J. Shotton, S. Hodges, A. Fitzgibbon, KinectFusion: Real-time dense surface mapping and tracking. *Proc ISMAR 2011* (2011), 127–136.

Address for correspondence

Corresponding author: Ekaterina Sirazitdinova,
E-mail: ekaterina.sirazitdinova@rwth-aachen.de.

A Document-Based EHR System That Controls the Disclosure of Clinical Documents Using an Access Control List File Based on the HL7 CDA Header

Toshihiro Takeda^a, Kanayo Ueda^b, Akito Nakagawa^a, Shirou Manabe^a,
Katsuki Okada^a, Naoki Mihara^a, Yasushi Matsumura^a

^a Department of Medical Informatics, Osaka University Graduate School of Medicine, Suita, Osaka, Japan

^b Division of Medical Informatics, Osaka University Hospital, Suita, Osaka, Japan

Abstract

Electronic health record (EHR) systems are necessary for the sharing of medical information between care delivery organizations (CDOs). We developed a document-based EHR system in which all of the PDF documents that are stored in our electronic medical record system can be disclosed to selected target CDOs. An access control list (ACL) file was designed based on the HL7 CDA header to manage the information that is disclosed.

Keywords:

Electronic Health Record, Information Management, Medical Records Systems.

Introduction

To facilitate the smooth continuation of medical treatment, it is necessary to share medical information between care delivery organizations (CDOs). Because electronic medical record (EMR) systems have been widely implemented in Japan, the data entered into the EMR database must be transferred to the electronic health record (EHR) database. Various standardized data can be shared via the Japanese standard repository, SS-MIX [1]. However, the existing EHR systems have two major problems: 1) it is still difficult to share non-standardized data, and 2) the EHR system does not offer detailed control of the documents that are to be disclosed. Our hospital implemented a document-based EMR system called DACS [2]. In this system, print images of all medical records generated by the EMR systems were stored in the DACS database in the PDF format. In this study, we demonstrate the document-based EHR system.

Methods

The existing EHR system was used for network security (OD-VPN), user management and patient management. The standardized data are opened using the existing EHR function. The non-standardized data can be opened from the DACS database as PDF documents. The doctor can disclose all clinical documents with the selection of the target CDOs and the documents to be disclosed. We designed an access control list (ACL) file based on the HL7 CDA header to manage the settings for document sharing. The ACL file includes document information (document ID, version, document class code, document title and event date), patient information (patient ID, patient name, sex, and birthday), and information about the original user and generator (facility ID, facility name, original user ID, original username, generator ID and

generator name) and the facility that is permitted to view the document (facility ID and facility name).

Results

The doctors in the service CDO select the target CDOs and the documents that they intended to disclose on the DACS viewer (Fig. 1). With this information, the DACS generates an ACL file and transfers selected PDF documents and their ACL file into the SS-MIX repository. The report viewer at the target CDO makes a list of the disclosed documents according to the information contained in the ACL file. The user in the target CDO can select the PDF images that they wish to browse.

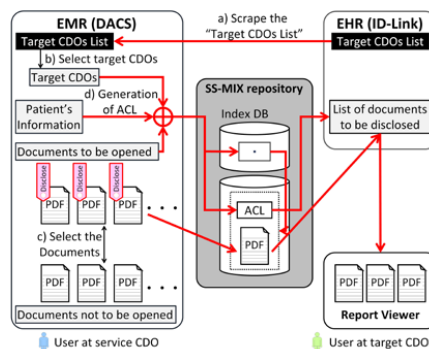


Figure 1- An overview of our document-based EHR system.

Conclusion

With the document-based EHR system, the documents are disclosed through the Japanese standard repository, SS-MIX and the document disclosure information is controlled by the ACL file, which is created based on the HL7 CDA header.

References

- [1] Kimura M, et al. SS-MIX: a ministry project to promote standardized healthcare information exchange, *Methods Inf Med.* 50 (2011), 131-9
- [2] Matsumura Y, et al. A scheme for assuring lifelong readability in computer based medical records. *Stud Health Technol Inform.* 160 (2010), 91-5.

Address for correspondence

Email: ttakeda@hp-info.med.osaka-u.ac.jp

School-Based Telemedicine: Perceptions About a Telemedicine Model of Care

May Lin Tye^a, Michelle Honey^b, Karen Day^a

^a Health Systems, School of Population Health, University of Auckland, Auckland, New Zealand,

^b School of Nursing, University of Auckland, Auckland, New Zealand

Abstract

A school-based store-and-forward telemedicine program, which is delivered to school children by school staff is an acceptable model of care for treating skin infections. This indicates that non-clinicians have potential to play a key role in telemedicine. Benefits include empowerment and role redefinition of staff, and potential improvements in health awareness of children, with no major issues perceived in the current program.

Keywords:

Telemedicine, school health services, staphylococcal skin infections

Introduction

Telemedicine is the delivery of medical care, involving an element of distance, and involves the use of information and communications technology (ICT). [1] Store-and-forward (SAF) telemedicine involves a collection of information which is sent to a physician who can respond in their own time. [2] In New Zealand (NZ), an SAF telemedicine program has been implemented in 21 schools in the Far North and in South Auckland as a potential solution and facilitator of access to care for children with skin infections. In NZ, the incidence of serious skin infections in children aged zero to 14 is high, with increasing hospitalizations which almost doubled between 1990 and 2007. [3] This program is delivered by school staff including teachers, principals, and administration staff. Within this telemedicine program, health information of children is collected and entered into an application (app) on a tablet or smartphone. This is sent to a clinical team who can remotely assess it and develop a management plan, which is authorized by a doctor. The research question is: How is a telemedicine model of care for skin infections in school children perceived by the school staff delivering the program? Sub-questions are: 1) How acceptable is telemedicine for school children to non-clinically trained people administering it? 2) What are some perceived benefits and issues with the current program?

Methods

This exploratory, descriptive study has been approached with qualitative methods, specifically an interpretivist approach. Sampling first involved sampling the schools, and secondly sampling staff within the schools. Schools were selected through convenience sampling. 13 schools were approached, with five agreeing to participate, and eight school staff being interviewed. A sociotechnical theory developed by Lamb and Kling reconceptualizes the user of ICT as a social actor. [4] The four constructs of this model: identities, affiliations, interactions, and environments, guided interviews, and analysis. A semi-structured interview, using an interview schedule was adopted in order to elicit narratives. Interviews were one-on-one, lasting up to one hour.

Results

Telemedicine aligned with *identities* of staff, fulfilling a duty to care for students, and fulfilling school values. *Affiliations* and close relationships with children, support from the community, and the telemedicine team enabled the program. Delivering telemedicine is an opportunity for staff to have *interactions* with children, and interactions between children reflected empowerment and awareness of health. *Environments*, in terms of embedded practices of the school, were influenced by the initial challenge for staff fitting telemedicine into workflows, and some concern of overuse of program by children. Physical environments were flexible, and staff recognized the importance of balancing the need for privacy while protecting a child's vulnerability.

Conclusion

School-based telemedicine delivered by school staff who are not clinicians is perceived as an acceptable model of care for skin infections in school children. Benefits include empowerment, potential improvement in health awareness, and literacy. No major issues perceived, except initial challenges to day-to-day practices, and balancing privacy and vulnerability. Non-clinicians can play a key role in telemedicine delivery.

Acknowledgements

This research was completed for a Bachelor of Health Sciences (Honors) dissertation at the University of Auckland. No conflict of interest exists.

References

- [1] Wootton, R. Telemedicine. *British Medical Journal* **323** (2001), 557.
- [2] Della Mea, V. Prerecorded telemedicine. *Journal of Telemedicine and Telecare* **11** (2005), 276-284.
- [3] O'Sullivan, C., Baker, M., & Zhang, J. Increasing hospitalizations for serious skin infections in New Zealand children, 1990–2007. *Epidemiology and Infection* **139** (2011), 1794-1804.
- [4] Lamb, R., & Kling, R. Reconceptualizing users as social actors in information systems research. *MIS Quarterly* **27** (2003), 197-236.

Address for correspondence

May Lin Tye - School of Population Health, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. mtye882@aucklanduni.ac.nz

CLARUS as a Cloud Security Framework: e-Health Use Case

David Vidal, Santiago Iriiso, Rafael Mulero

Information System Department, Fundació Clínic per a la Recerca Biomèdica (Hospital Clínic i Provincial de Barcelona),
Barcelona, Spain

Abstract

Maintaining Passive Medical Health Records (PMHR) is an increasing cost and resource consumption problem. Moving to the cloud is the clearest solution to solve the problem as it offers a high amount of space and computation power. But the cloud is not safe enough when dealing with this kind of information because it can be easily accessed by attackers. The European Commission funded research project CLARUS contributes to protect healthcare-sensitive information in a secure way.

Keywords:

Telemedicine, Medical Health Records

Introduction

The PMHR is the information from patients who have no more contact with the hospital in a time period of 5 or more years for any reason (moving to another country, death). In some cases, such as patient healthcare, research or legal requirements, this information has to be accessible to professionals from the hospital as well as external healthcare professionals if needed.

Nowadays, this data is being stored on premises where space is limited and the resources for computing this amount of data are not enough as they are being used to manage the non-passive medical health records. Thus, moving to a cloud solution looks like the clearest solution to store and process this information providing more space to local servers and decreasing the resource consumption.

The main problem is that the e-health data privacy and security laws [1] are very strong. In this context, it's therefore necessary to apply a solution ensuring the security and confidentiality of the information in order to use a cloud solution not only to store but also to compute the data. Cloud solutions nowadays don't offer this level of security for this kind of information which is one of the main reasons that they are not being extensively used in the healthcare sector. CLARUS [2] is a Horizon 2020 European Commission funded project where the main objective is to design a framework to protect sensitive information in the cloud.

Methods

The Information System Department (ISD) of the Hospital Clínic de Barcelona has collaborated in the CLARUS project as a use case where a framework is applied to healthcare data. First, the ISD elaborated a synthetic dataset based on real distribution of the data, in order to have a fake dataset to work with. The ISD also created a demonstrator application showing

the process of retrieval and computation of this data using the CLARUS solution. Encryption (searchable- encryption) or anonymization (k-anonymity and t-closeness) was applied to add more security to the dataset before moving it to the cloud. With searchable-encryption, the data is encrypted on cloud, making it unreadable, and it's possible to make queries directly over the encrypted dataset without prior decryption. With k-anonymity and t-closeness the data is masked or changed in some parts in order to preserve the privacy and usability of the data.

Results

In the case of data retrieval using the CLARUS solution:

- The information accessed directly in the cloud is completely useless for unauthorized users.
- Using CLARUS, the encrypted information is returned in clear for allowed users.

In the case of the data cloud computation:

- The information in the cloud doesn't show any identifier or data that could identify a patient.
- When comparing the computations over the original and anonymized dataset, a non-significant deviation appears depending on the filters applied for the computation queries.

Conclusion

Using CLARUS as a framework for cloud security allows the storage, retrieval and in cloud computing of e-health sensitive information with a high level of protection. To ensure the usability of this solution more tests are needed, not only to improve the cloud computation, but also to avoid any access attempts by unauthorized users.

References

- [1] Spanish Data Protection Agency (2014) Agencia Española de Protección de Datos. Retrieved November 24th, 2016, from <http://www.agpd.es/portalwebAGPD/internacional/index-ides-idphp.php>
- [2] Clarus (2016) CLARUS. Retrieved November 24th, 2016, from <http://clarussecure.eu>
- [3] L. Sweeney, k-anonymity: a model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10** (2002), 557-570.
- [4] N. Li, T. Li and S. Venkatasubramanian (2007), t- Closeness: Privacy Beyond k-Anonymity and l-Diversity. IEEE 23rd International Conference on Data Engineering.

Address for correspondence

Santiago Iriiso, siriso@clinic.cat

An Interdisciplinary Approach Between Medical Informatics and Social Sciences to Transdisciplinary Requirements Engineering for an Integrated Care Setting

Jan Vielhauer^a, Britta Böckmann^{a,b}

^a Department of Medical Informatics, University of Applied Sciences, Dortmund, NRW, Germany

^b IMIBE, University Hospital Essen, Germany

Abstract

Requirements engineering of software products for elderly people faces some special challenges to ensure a maximum of user acceptance. Within the scope of a research project, a web-based platform and a mobile app are approached to enable people to live in their own home as long as possible. This paper is about a developed method of interdisciplinary requirements engineering by a team of social scientists in cooperation with computer scientists.

Keywords:

Mobile Applications, Health Services Needs and Demand, Home Care Services

Introduction

Due to the demographic change in society, a growing number of people will live to an older age and thus the demand for nursing and care will increase [1,2]. In this project, we will help to reveal and capitalize resources supporting the patient's daily activities and delay the need for inpatient care for as long as possible by means of a digital networking platform. This work is about a new method to gather and control the requirements of very specific and heterogeneous actors by combining methods from social science with scrum.

Methods

Considering the heterogeneity of the actors in this process, a team of software engineers, medical information scientists, and social scientists has been formed who will conduct the requirements engineering and reveal the special needs and potential resources. To get to a point where medical computer scientists and social scientists working together on the derivation of requirements the working methods of the domains need to be merged. In a setting where software engineers are using scrum as an agile and iterative development method, the requirements may be changed at any time. The results of the feedback from the scrum product owner are to be reevaluated and then stored as new items in the product backlog [4]. Requirements engineering from the social scientist teams point of view is done by open interviews with a snowball sampling technique [3]. The interviewed persons will be asked to provide more persons that could be relevant as a requirement source for the project and thus build and extend a focus group. Members of this group will be interviewed in the next iteration step with the improved interview guideline of the last iteration step. In order to get to a point where the software engineers can use the gathered information to build a software, Scrum User Stories became apparent as a suitable method of extracting concrete

requirements out of ideas and impressions by non-computer-scientists [4].

Results

As a result, a three-step iterative way of gathering and detecting requirements has been developed. It will put out and reevaluate user stories to the software engineers building the software. In the first step, a nursing service is accompanied in their daily work by a social and a medical computer scientist. Ideas for improvement will be gathered passively. The second step is a comparison of information by the engineers which aims to reveal missing information and review ideas by both experts so that unusable ideas can be rejected out of hand. The final step is to interview the focus group members and generate improved user stories with every iteration step.

Conclusion

First results of the developed method have shown it to be very effective in view of the aim of the project, to reevaluate both the social and the technical requirements iteratively. The resulting requirements from a both points of view are wide-ranging, yet goal-driven, as the developed method creates needed transparency for all actors in the development process. Future research will show how effective the developed software from our requirements will be regarding user acceptance and prolonged time in the living environment of the patient.

References

- [1] A. Brylok, A. Karmann, S. Becker, R. Schneider, U. Zimmermann, Städtebau der Zukunft. Selbstständiges Wohnen bis ins hohe Alter – eine volkswirtschaftliche Analyse. *Eigenverlag, Dresden* (2015).
- [2] H. Rothgang, Umverteilung von unten nach oben. *Soziale Sicherheit, Der "Pflege-Bahr"* (2012) Heft 6: 204.
- [3] C. Noy, Sampling Knowledge: The Hermeneutics of Snowball Sampling in Qualitative Research. *International Journal of Social Research Methodology* (2008), 327-344.
- [4] M. Cohn, User Stories Applied: For Agile Software Development. *Addison-Wesley Professional, Boston* (2004).

Address for correspondence

Jan Vielhauer, University of Applied Sciences and Arts, Department of Medical Informatics, Emil-Figge-Straße 42, 44227 Dortmund, Germany, jan.vielhauer@fh-dortmund.de

Use Contexts and Usage Patterns of Interactive Case Simulation Tools by HIV Healthcare Providers in a Statewide Online Clinical Education Program

Dongwen Wang, PhD

Arizona State University, Scottsdale, AZ, USA

Abstract

We analyzed four interactive case simulation tools (ICSTs) from a statewide online clinical education program. Results have shown that ICSTs are increasingly used by HIV healthcare providers. Smart phone has become the primary usage platform for specific ICSTs. Usage patterns depend on particular ICST modules, usage stages, and use contexts. Future design of ICSTs should consider these usage patterns for more effective dissemination of clinical evidence to healthcare providers.

Keywords:

Computer Simulation, Information Dissemination, HIV

Introduction

In recent years HIV research has made rapid progress. Timely dissemination of the latest findings to healthcare providers is a key step for improvement of HIV patient care. We previously reported the development of interactive case simulation tools (ICSTs) for dissemination of HIV clinical evidence and the initial usage data of an ICST for *Insomnia Screening and Treatment* [1]. In this study, we further analyze ICST usage by including the use contexts, extending the study to additional ICST modules, and conducting assessment in both the initial and stable usage stages. These analyses will identify ICST usage patterns and assist the design of effective approaches for dissemination of clinical evidence through ICSTs.

Methods

The ICSTs in this study were developed for New York State (NYS) HIV-HCV-STD Clinical Education Initiative (CEI). We selected four ICSTs for analyses: (1) *Insomnia Screening and Treatment*; (2) *Mental Health Screening*; (3) *Post-Exposure Prophylaxis Following Occupational Exposure (oPEP)*; and (4) *PEP Following Sexual Assault (nPEP)*. Development of these ICSTs was based on the related clinical guidelines. The initial usage stage was 195 days for the *Insomnia ICST* and 97 days for the other three ICSTs. The stable usage stage for the *Insomnia* and *Mental Health Screening* ICSTs was 111 days. For data analyses, we characterized ICST use context by: (1) *new vs. returning user*; (2) *access from large-screen equipment vs. small-screen hand-held device*; and (3) *use through web browser vs. native app*. These use contexts were extracted from the system logs of the CEI web servers. To analyze ICST usage patterns, we profiled the frequency of visits to specific ICST sections, i.e., *recommendation*, *sample case*, *user-defined case*, and *cross-box*, by each dimension of the use contexts described above. For each of these usage profiles, we used chi-square test to examine the statistical significance.

Results

In the initial usage stage, we recorded a total of 512 visits to the *Insomnia ICST*, 422 visits to the *Mental Health Screening ICST*, 82 visits to the *oPEP ICST*, and 80 visits to the *nPEP ICST*. Analyses of use contexts and ICST sections found 82-98% visits from *new users*, 48-78% visits from *native apps*, and 34-70% visits from *large-screen equipment*. Usage pattern analyses found that: (1) *new users* were more likely from *small-screen devices*; (2) visits to *user-defined cases* were more likely from *small-screen devices*; and (3) *small-screen users* were more likely to access ICSTs through *native apps*.

In the stable usage stage, we recorded a total of 898 visits to the *Insomnia ICST* and 1560 visits to the *Mental Health Screening ICST*. Analyses of use contexts found 79-92% visits from *new users*, 81-90% visits from *native apps*, and 26-31% visits from *large-screen equipment*. Usage pattern analyses found that *new users* were more likely to use *large-screen equipment*.

Conclusion

We have characterized the use contexts of four ICSTs in both the initial and stable usage stages, and identified specific usage patterns and their changes in the two usage stages. Results from these analyses have shown that: (1) ICSTs are increasingly used by HIV healthcare providers; (2) smart phone has become the primary usage platform for ICSTs; and (3) various other usage patterns are related to particular ICST modules, usage stages, and use contexts. Future design of ICSTs should consider these use contexts and usage patterns for more effective dissemination of clinical evidence to healthcare providers.

Acknowledgements

This work is supported by grant #R24HS022057 from the Agency for Healthcare Research and Quality (AHRQ), and by contracts #C023557, #C024882, and #C029086 from NYS Department of Health AIDS Institute.

References

- [1] Le XH, Luque A, Wang D. Assessing the usage of a guideline-driven interactive case simulation tool for insomnia screening and treatment in an HIV clinical education program. *Stud Health Technol Inform.* 2013;192:323-7.

Address for correspondence

Dongwen Wang, PhD; Professor of Biomedical Informatics, Arizona State University; 13212 E. Shea Blvd., Scottsdale, AZ 85259; dongwen.wang@asu.edu; 1-480-884-0264.

A Homemade Easy Information System: The Example of Patient Satisfaction Survey

Hang Yin^{ab}, Pingping Yan^a, Yanyan Cui^c, Polun Chang^c, Jun Lei^a

^a The Third Xiangya Hospital of Central South University, Changsha, Hunan, China

^b Clinical Medical College of Tianjin Medical University, Tianjin, China

^c Institute of BioMedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

Abstract

An informatics system was developed based-on Excel Visual Basic for Applications (Excel VBA). It allows patients waiting for appointment complete their admission or satisfaction forms on a Window-based mobile device in-stead of paper. Their data will be stored in the tool and analytical charts will automatically be presented in the interfaces.

Keywords:

Information systems; Students, Nursing; Appointments and schedules

Introduction

Patients usually were required to complete some forms for gathering information during the first clinic visit. Paper questionnaires as a traditional way were most frequently used if necessary. However, using questionnaires not only has a potential problem for long-term storage, but forces nurses to allocate time to enter the information into a database. Research shows nurses enabled with information technology would reduce indirect care time and improve patients' satisfaction [1]. We designed an information system based on Excel VBA to make it easy to collect data, automate data analysis and present our results.

Methods

The system was designed at the end of the Excel VBA Courses for masters in nursing informatics [2]. The core technical components of the system was composed of module design, creation of survey data pivot tables and animated charts from a database or list enable nurses to summarize data. The questionnaire was designed based on Marshall's patient satisfaction questionnaire [3], which consisted patient's information and their satisfaction about nursing service.

Results

Our system started with a friendly splash screen and the login interface. The option to validate the data in an automated way was available. Data could automatically be summarized, analyzed, and presented by flexible pivot tables, charts, and graphs (Fig.1). It is transparent to demonstrate which nursing service needed to be improved. All data were stored in electronic form and can be exported to another statistical software.



Figure 1 - The informatics system screenshot displaying that all data could automatically be summarized, analyzed, and presented by charts and be stored in electronic form

Discussion

Using paperless and electronic scales on mobile platforms for patient check-in or information collection in hospital would drive trend in the future. The system is a practical and user friendly tool that allows users to select by touching the screen with their fingertips. As a functional platform, it also allows for data pooling and automatically analyses.

Informatics competency is an essence skill for all nurses and nursing students. A VBA training program appeared to be a practical strategy to promote nursing informatics by allows nurses to design susyem on their own.

Acknowledgements

Funding was provided by the New Xiangya Talent Project of the Third Xiangya Hospital of Central South University (Zhishanlingpao Program:20170305).

References

- [1] I.C. Hou, C.W. Liu, C.L. Hsu and et al, The evaluation of using road-map guided Excel VBA training to promote nursing informatics in Taiwan[C]//Nursing Informatics (2009), 253-257.
- [2] P. Chang, M.C. Kuo, Taiwan Model: Nursing Informatics Training. In Ball MJ, Hannah KJ, Newbold SK, Douglas JV. (ed.) Nursing Informatics: Where Caring and Technologies Meet. (4th edition). New York, NY, USA: Springer.2010:411-28.
- [3] G.N. Marshall, and R.D. Hays, The patient satisfaction questionnaire short-form (PSQ-18) [J] (1994), 7865.

Address for correspondence

Jun Lei - leijunbao@126.com

Personal Health Self-Management in a Data Perspective

Yiqin YU^a, Xiang LI^a, Gang HU^a, Haifeng LIU^a, Jing MEI^a,
Yuan Ni^a, Guoyu Tang^a, Guotong XIE^a, Weiming XU^b

^aIBM Research – China, Beijing, China, ^bIBM Global Business Services, Wuhan, China

Abstract

Along with the growth of numbers of patients with chronic diseases, personal health self-management becomes critical. The heterogeneity of self-management requirements makes the detail design and implementation of self-management program a non-trivial work. In this paper we address the problem with the Personal Health Advisor (PHA) application by introducing a personal health data flow mechanism, as well as modules including personal health risk assessment, similar patients profiling, and health question answering.

Keywords:

Self Care, Risk Assessment, Health Education

Introduction

While evidence suggests that chronic disease self-management programs can improve health status and outcomes while reducing hospital visits [1], the heterogeneity of self-management requirements makes the implementation of self-management tools a non-trivial task. In this paper, we try to address the personalized health self-management problem in a mobile-based, data-driven and patient centric mode. We design the modules following a “cognitive” process for the self-management of chronic diseases, that is: awareness of the current health status and potential health risks, comparison with similar patients, and learn from questioning and answering.

Methods

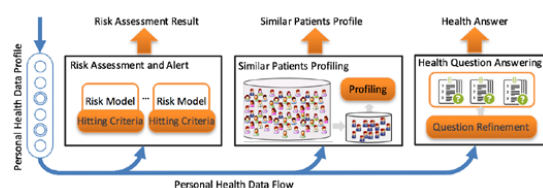


Figure 1. Personal Health Data Flow Mechanism

The Personal Health Advisor (PHA) is implemented as a mobile app. All the components work with the personal health data flow mechanism illustrated in Figure 1. The Printed Medical Records Extractor module is used to collect and extract structured clinical data from uploaded medical record photographs, whose precision reaches 0.945 and recall 0.880 [2]. Each input will generate a new instance of the Personal Health Data Profile. Then, the data profile instance will be passed to the Risk Assessment and Alert module. Once the hitting criteria of a certain model in this module is triggered, the corresponding model is applied to the current data profile and generating health risk scores, risk factor impacts, and risk alerts. The Similar Patients Profiling module calculates the distance

among the current profile and profile of other users, and generates health summaries using the most similar users. With the current profile, the user can also ask oral questions about his/her health conditions and get answers from the Health Question and Answering module, which are retrieved from online public Q&A service websites in China.

Case Study

Two case studies were performed. In the first one, the PHA was used for the personal health self-management of Chinese elder patients with chronic conditions. By initializing the specified disease suit for elder patients, PHA adapts itself to the generic progression of elder users' health conditions and provides up-to-date risk assessment results for 10-years stroke onset and 4-years congestive heart failure onset, summarizing 2000 similar patient's profiles who might also have diabetes and hypertension, and personalized educational health information. In another case, in cooperation with the obstetrical department of a private hospital in China, we integrated PHA with their register appointment app to enhance self-management. The disease suit is specified for common pregnancy diseases in China including gestational diabetes mellitus and preeclampsia. Once an agreement is signed, the registered pregnant women can visit their health record from this hospital directly from the app, or can upload medical record photograph from other hospitals. Personal risks for the two diseases are calculated. Answers for pregnancy knowledge and diet and exercise-related questions during pregnancy are also provided.

Conclusion

PHA is a practical tool for personal health self-management. It facilitates user centric health self-management by applying health data collection, risk assessment, patient similarity analysis, and question answering in a “cognitive” way.

References

- [1] K.R. Lorig, D.S. Sobel, A.L. Stewart, B.W. Brown Jr, A. Bandura, P. Ritter, V.M. Gonzalez, D.D. Laurent, H.R. Holman, Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. *Med Care* **37** (1999), 5-14.
- [2] L. Xiang, G. Hu, T. Xiaofei, G. Xie, Building structured personal health records from photographs of printed medical records. *AMIA Annual Symp Proc* **2015** (2015).

Address for correspondence

Yiqin Yu, MS. IBM Research – China, Diamond Building, Haidian District, Beijing, China, 100193. Email: yuyiqin@cn.ibm.com

Nurses' Interest, Readiness and Absorptive Capacity to Information Technology: A Survey in China

Boning Zhang^a, Ping Liu^b, Qian Xiao^c

^aBeijing Children's Hospital, Capital Medical University, Beijing, China

^bXuanwu Hospital, Capital Medical University, Beijing, China

^cSchool of Nursing, Capital Medical University, Beijing, China

Abstract

To investigate nurses' interest, readiness and absorptive capacity to information technology, 261 nurses were investigated using anonymous questionnaires. This study showed: 1) the top 3 information technologies were personal digital assistant (PDA), Hospital Information System (HIS) and wireless mobile nursing trolley; and 2) the mean scores of interest, readiness and absorptive capacity to information technology were 16.3, 56.7 and 46.8, respectively. Further educational programs should be provided for nurses.

Keywords:

Capacity Building; Nursing Informatics; Education

Introduction

Information technology (IT) has been increasingly available for health care and nursing work, which has not only changed the traditional mode of nursing, but also promoted disease management and the efficiency and quality of nursing. Technology readiness refers to individuals' overall state to embrace and adopt new technology. [1,2] Absorptive capacity represents individuals' external technology acquisition activity and innovation performance. Given the increasing use of information technologies in nursing, this study aimed to explore the technology interest, readiness and absorptive capacity of nurses in China.

Methods

In hospitals in China, 261 nurses with more than 6 months working experience were surveyed using a questionnaire to perform a cross-sectional survey. The questionnaire mainly includes four aspects: (1) general information, including age, education, working period, and position, (2) interest in information technology, (3) readiness to use information technology, assessed based on the Parasuraman technology readiness index [4], and (4) absorptive capacity towards information technology, assessed based on an IT absorption capacity scale. [3]

Results

The top 3 information technologies used by nurses were PDA, HIS and wireless mobile nursing trolleys. The score of nurses' interest in using information technology ranged from 4 to 20, with a mean score of 16.3 (SD 2.8). Different education level and position at work are associated with interest score, with statistical significance ($P < 0.05$). The average technology readiness score was 56.7 (out of 90) and positively correlated with the interest score ($P < 0.01$).

The average score of nurses' absorptive capacity to technology was 46.8 (SD 7.5), which was positively correlated with interest score and readiness ($P < 0.01$).

Discussion

Higher interest and higher readiness appear to be associated with higher overall information technology absorptive capacity. Interest in IT and readiness to use IT are both important to enhance the technology absorptive capacity of nurses.

Conclusion

Continuous educational programs should be provided for nurses to enhance their interest, readiness and absorptive capacity to information technology.

References

- [1] Kuang-Ming Kuo, Chung-Feng Liu, Chen-Chung Ma. An investigation of the effect of nurses' technology readiness on the acceptance of mobile electronic medical record systems, *BMC Med Inform Decis* **13** (2013), 88-91.
- [2] Amy L. Caison, Donna Bulman, Shweta Pai and Doreen Neville. Exploring the technology readiness of nursing and medical students at a Canadian University, *J Interprof Care* **22** (2008), 283-294.
- [3] Yu Cuiling, Chen Taibo, Shi Ying. Information technology absorption capacity: the mechanism research of information system success, *Journal of the China Society for Scientific and Technical Information* **31** (2012), 748-758 (Chinese).
- [4] Parasuraman a. Technology readiness index (tri): A multiple-item scale to measure readiness to embrace new technologies, *J Serv Res* **2** (2000) and **4** (2000): 307-320.

Address for correspondence

Qian Xiao, Associate Professor.

School of Nursing, Capital Medical University, Beijing, China.

Email: julia.xiao@163.com, Phone: 86-10-83916507

This page intentionally left blank

II. Health Data Science

This page intentionally left blank

A Deep Learning Approach to Neuroanatomical Characterisation of Alzheimer's Disease

Abhinit Kumar Ambastha^a, Tze-Yun Leong^{a, b}, The Alzheimer's Disease Neuroimaging Initiative

^a Medical Computing Laboratory, School of Computing, National University of Singapore, Singapore

^b School of Information Systems, Singapore Management University, Singapore

Abstract

Alzheimer's disease (AD) is a neurological degenerative disorder that leads to progressive mental deterioration. This work introduces a computational approach to improve our understanding of the progression of AD. We use ensemble learning methods and deep neural networks to identify salient structural correlations among brain regions that degenerate together in AD; this provides an understanding of how AD progresses in the brain. The proposed technique has a classification accuracy of 81.79% for AD against healthy subjects using a single modality imaging dataset.

Keywords:

Alzheimer Disease; Machine Learning; Artificial Intelligence

Introduction

Alzheimer's disease (AD) is a chronic degenerative disease that occurs in middle or old age, due to neuronal loss in the brain. It is a prevalent cause of premature senility and leads to a steady decline in a person's capability to carry out activities of daily living. Early detection of AD enables preventive measures and lifestyle changes to be employed to slow down the progression of the disease. Current evidence supporting AD management are mainly based on epidemiological studies. Existing studies that involve computational methods usually provide effective classification but minimal explainability. For example, Heung-II Suk et al [1] conducted multimodal fusion with deep learning for AD/Mild Cognitive Impairment classification; Adrien Payan and Giovanni Montana [2] used deep convolutional neural networks to analyse and classify AD patients from healthy subjects. Neuroanatomical characterisation of AD using a computational method is a novel approach. From an anatomical perspective, atrophy and connectivity changes in AD are well studied topics, and it is established that these changes can be used to explain behavioural changes in AD subjects. In this work, we focus on atrophy-based neuroanatomical characterisation of AD.

Methods

We propose a method to quantify the correlated regional atrophy that occurs in the brain during the course of AD progression. We examine a set of cross-sectional structural magnetic resonance images (MRI) from The Alzheimer's disease Neuroimaging Initiative (ADNI) data repository in this study. We use MRI data from 100 patients in the screening stage of the ADNI-1 study (50 AD, 50 Healthy Controls), between the ages of 59 and 88. These images were preprocessed, normalised and registered to a template. We extract and select voxel features from the brain regions using

deep neural networks and identify their contributions towards AD using an ensemble learning algorithm. We use convolutional neural networks with dual region inputs to detect correlated degeneration between the input regions and apply the AdaBoost algorithm to identify the saliency of the *region pairs* towards AD. We construct brain region networks by detecting the edges among brain regions that degenerate together in AD.

Results

The output of our method is an auto-associative model which identifies cliques of brain region networks that degenerate together in AD. We compared our work with classification using Bayesian structure learning algorithms. The classification accuracy of our model is 81.79% compared to 74.01% for score-based structure learning algorithms and 74.42% for constraint-based structure learning algorithms.

Conclusion

The identified regional pairs can potentially explain several behavioural changes that commonly occur in AD patients, e.g., the cerebellar vermis region was found to atrophy in a correlated manner with the left hippocampus region. Cerebellar vermis is the region responsible for locomotion and gait, which is observed to be disturbed in AD patients. These correlations are aligned with other statistical and computational study findings which motivated this work [1,2]. Our future plans include automatically selecting regions of interest (ROI) based on a region proposal network (RPN). Using an RPN will remove the dependency on the atlas selected for sampling the brain regions. The next stage is to identify the neuronal tracts in the brain which degenerate together. Neuronal tracts enable inter-region signaling in the brain. Examining the damage progression of these tracts will help predict the behavioural changes in AD patients.

References

- [1] H.-I. Suk and D. Shen, Deep learning-based feature representation for AD/MCI classification, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2013, pp. 583-590.
- [2] A. Payan and G. Montana, Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks, *arXiv preprint arXiv:1502.02506* (2015).

Address for correspondence

Abhinit Kumar Ambastha, email : abhinit@comp.nus.edu.sg

Three Dimensional Virtual Planning Through Cone Beam Computed Tomography for Surgical Guidance Production

Kaline Thumé Antunes^a, Marcos Cordeiro d'Ornellas^{a,b},
Daniani Deliberalli Noedel^a, Gustavo Nogara Dotto^{a,b}, Renésio Armino Ghres^b

^aGrupo de Computação Aplicada em Saúde (CA+SA), Universidade Federal de Santa Maria, Santa-Maria-RS, Brasil

^bHospital Universitário de Santa Maria (HUSM), Universidade Federal de Santa Maria, Santa Maria-RS, Brasil

Abstract

This paper describes a methodology through digital planning and computed tomography for the purpose of making surgical guides for insertion of skeletal anchorage devices in dentistry. Mini-orthodontic screws for anchorage are inserted virtually in the region of interest located between the teeth of the second premolar and first molar bilaterally. DICOM images were converted to STL format and overlaid by scanning the upper plaster model in order to plan the surgical guide virtually.

Keywords:

Radiography, Panoramic; Cone-Beam Computed-Tomography; Radiation Dosage

Introduction

Biomodels, replicas of an anatomical structure, are produced through a chain of processes that begins with the acquisition of computed tomography (CT) and magnetic resonance imaging (MRI). DICOM files are preprocessed using specific applications, which allow to generate a 3D representation of the surface of the object. The resulting file has an STL extension (stereolithography) that is used to manufacture the biomodel.

Considering the increasing applicability of Computerized Tomography in Dentistry, it is very important to determine which image acquisition protocol is capable of providing a three-dimensional visualization with resolution and sharpness appropriate for measuring the structures.

Methods

Computed tomography were obtained with voxel of 0.25mm, time of acquisition of 20s. Plaster models were also digitized in the same way with DICOM 3.0 format archiving for 3D STL reconstruction in appropriate software. The anchoring is decisive for the success of the orthodontic treatment, existing several intra and extraoral resources to be used. After this preparation of the images and virtual planning, the mini-implant is created in the program and its properties such as diameter and thickness are determined. The digital model is used to select the region on which the guide will be drawn (Fig. 1).

The software calculates the characteristics of the design so that the guide can be milled with the use of biocompatible and transparent material (Acrylic Resin - VIPI). Mini-implants can contribute to a stable intraoral bone anchorage, without risk of reciprocal movement of the active unit as a reaction to the

action of orthodontic forces generated in the anchorage unit, and may be installed in several areas of the alveolar bone.

Results

The results have shown countless advantages. The degree of stability, better exploration of the available locations for the devices to be inserted, and the integrity of adjacent structures are directly influenced when a technique is performed. However, the proximity of the dental roots may compromise the stability of the mini-implant and thus narrower regions require very well-planned and accurate techniques [1]. The images generated by a Tomograph have been improved, seeking not only the gain in image quality and the reduction of artifacts, but also the reduction of the radiation absorbed by the patient during the examination.

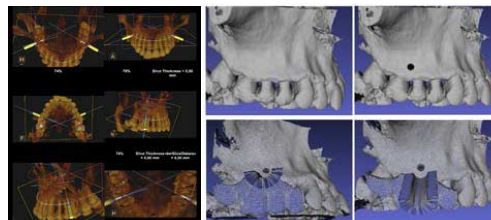


Figure 1 - 3D reconstruction of maxillary teeth and bone, virtual marking of the positioning of mini-implants and their angulation and virtual planning of the surgical guide.

Conclusion

A new approach for making surgical guides to insert mini-implants has been presented. Through computerized tomography and digital models, the guide made using dental milling has the potential to reduce clinical interferences, such as root damage or perforation of other anatomical structures.

References

- [1] S. Kuroda, K. Yamada, T. Deguchi, T. Hashimoto, H.-M. Kyung, and T.T. Yamamoto, Root proximity is a major factor for screw failure in orthodontic anchorage, *American Journal of Orthodontics and Dentofacial Orthopedics* 131 (2007), S68-S73

Address for correspondence

Kaline Thumé Antunes, kalinethantunes@gmail.com, Av. Roraima, 1000, Prédio 07, Anexo B, Sala 388 UFSM- Campus Universitário 97105-900 Santa Maria-RS – Brasil.

“Big Clinical Data” Analysis of Intravenous to Oral Conversions of Non-Antimicrobials

Patrick E. Beeler^{a,b}, David W. Bates^b, Jürg Blaser^a

^a Research Center for Medical Informatics, University Hospital Zurich & University of Zurich, Switzerland,

^b Division of General Internal Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, USA

Abstract

Advantages of early intravenous (IV) to oral (PO) switches of antimicrobials are well known. Yet, little data have been published on reasonable IV-to-PO switches of other drug classes, although interventions to promote early IV-to-PO conversions may further increase patient safety and decrease costs. We therefore analyzed IV-to-PO switches of non-antimicrobials.

Keywords:

Administration, Intravenous; Pharmacoepidemiology

Introduction

Early intravenous (IV) to oral (PO) conversions may increase patient safety and decrease costs. Studies successfully promoted earlier IV-to-PO conversions of antimicrobials [1]. Yet, little data have been published on IV-to-PO conversions of other drug classes. This analysis aims to generate more knowledge on IV-to-PO conversions beyond antimicrobials.

Methods

In this retrospective study at a Swiss teaching hospital (850 beds) all orders for inpatients between 8/2009-4/2014 were processed. When an IV drug was discontinued, the time frame -0.5 to 3.0 hours (interquartile range, IQR) was used to find starting PO therapies. The approach how we calculated the IQR is explained in figure 1. However, not only IV therapies that were switched to PO in the process were analyzed but also IV therapies without switch, i.e. therapies either discontinued or the patient was discharged. Only throughout the hospital stay were durations of IV and PO administrations considered.

Results

A total of 547,153 IV orders were analyzed. Of them 69,981 (12.8%) were switched to PO. Among frequently ordered IV drugs (table 1) metamizole was switched to PO in 37.4% of cases, whereas potassium was switched in only 6.0%. For table 2, only drugs were considered that were switched to PO ≥ 30 times and had mean IV administration durations of ≥ 3 days: those compared to IV drugs not switched in the process, the latter had up to 3 days shorter IV administration durations which may suggest potential for earlier switches as well.

Conclusion

“Big clinical data” analyses are an important pharmacoepidemiological tool to identify targets for interventions in order to improve patient safety and cost efficiency of drug therapies.

Table 1 – Mean IV and PO durations in frequent switches

IV drug	IV duration [days]	PO drug	PO duration [days]	Count
Paracetamol	1.13	Paracetamol	2.63	16416
Metamizole	1.25	Metamizole	2.64	12807
Esomeprazole	3.74	Esomeprazole	5.62	3837
Metamizole	1.02	Paracetamol	2.5	3746
Potassium	1.38	Potassium	3.07	2623

Table 2 – Mean IV duration with vs. without following switch

IV drug	IV duration [days] (later switched to PO)	IV duration [days] (not switched)	Difference [days]
Mycophenolic acid	8.62	5.58	-3.04
Methylprednisolone	3.3	1.19	-2.11
Tranexamic acid	3.48	2.43	-1.05
Pantoprazole	4.3	3.51	-0.79
Esomeprazole	3.71	3.13	-0.58

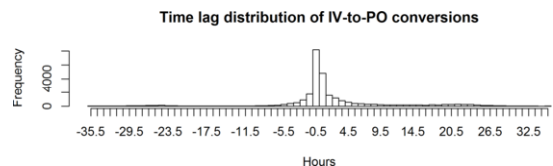


Figure 1 – At the end of an IV therapy the time frame ± 36 hours was used to search for beginning PO administrations of the same drug agent. Thereby we found that half of PO drugs started within -0.5 h to +3.0 h after IV discontinuation

Acknowledgements

The authors thank Michael Fetzter for extracting the data.

References

- [1] P.E. Beeler, S.P. Kuster, E. Eschmann, R. Weber and J. Blaser, Earlier switching from intravenous to oral antibiotics owing to electronic reminders. *Int J Antimicrob Agents* **46** (2015), 428-433.

Address for correspondence

Patrick E. Beeler, patrick.beeler@usz.ch

MetaMap Lite in Excel: Biomedical Named-Entity Recognition for Non-Technical Users

Ravi Teja Bhupatiraju, Kin Wah Fung, Olivier Bodenreider

National Library of Medicine, Bethesda, MD, USA

Abstract

We developed an easy-to-use tool for non-technical biomedical researchers to conduct Named-Entity Recognition (NER) on biomedical text, in a familiar spreadsheet environment. The system is a simple, offline, easy to install, end-user front-end to the new MetaMap Lite. Early adopters found it to be a quick start-point to incorporate NER in their investigations.

Keywords:

Natural language processing; Unified medical language system.

Introduction

The application of Named Entity Recognition (NER) has become pervasive. Biomedical researchers, who may not have strong computer skills, often wish to apply NER methods and tools to extract information from text.

MetaMap (<https://metamap.nlm.nih.gov/>) is one of the most popular tools for biomedical Named Entity Recognition (NER), more specifically for identifying terms from the Unified Medical Language System (UMLS) Metathesaurus in biomedical text. MetaMap Lite is a recent Java reimplementation of the original MetaMap. Running these tools on biomedical text and parsing their output generally requires some programming skills, which places them out of reach for non-technical users. Our objective is to make biomedical NER tools easier to use by non-technical users.

Methods

We developed an easy-to-use tool for non-technical biomedical researchers to use MetaMap Lite on biomedical text, in a familiar spreadsheet environment, supporting interactive and batch processing operations.

Our system does not depend on network or external resources. Instead, a zero-configuration backend server provides an HTTP service that a spreadsheet function consumes to perform named

entity recognition. The function supports output field selection (e.g., "pref,type" returns the preferred name and semantic type, along with the UMLS concept unique identifier, or CUI). Matched text and source vocabulary may also be requested. By default, the system only returns the UMLS CUI and the preferred name). Semantic type restriction may also be specified (e.g., "phsu,antb" returns only those terms that have been categorized as pharmaceutical substances or antibiotics).

The backend server serves a self-documenting spreadsheet template for users to get started. It supports automatic update of NER results as users edit entries, and batch processing by dragging fill handles to apply the function to rows of natural text inputs. The function may be combined with other functions for further automation.

Results

Figure 1 illustrates a typical use case for our tool. Users copy biomedical text in one column (A) and use the *mmlite* function in another column (B) to identify UMLS concepts from the text in column (A).

From a technical perspective, the *backend* can run anywhere a Java Virtual Machine (JVM) is available. The Windows installer for the software package contains all the necessary software components for running the *mmlite* function in Excel.

On informal inquiry, users found the software easy to install and use. Response times were quick, at about 30ms per request on a Xeon E5-1620 v3 3.5 GHz with 16 GB RAM.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

Address for correspondence

Olivier Bodenreider: olivier@nlm.nih.gov

	A	B
1	Do not coadminister aliskiren with Exforge in patients with diabetes see Drug Interac	C1120110 aliskiren, C1949333 Exforge
2	Correction of electrolytes by dialysis may precipitate signs and symptoms of digitalis	C0013832 Electrolytes, C0304520 Digitalis preparation
3	Significant losses of protein, amino acids and water soluble vitamins may occur durin	C0002520 Amino Acids, C0259754 Water soluble vitamin product
4	Exforge may be administered with other antihypertensive agents.	C1949333 Exforge, C0003364 Antihypertensive Agents
5		

Figure 1 - Example of use of the *mmlite* function in conjunction with fill handles in quickly applying NER

An Exploratory Analysis of Questions Submitted to a Brazilian Telemedicine System

Jacqueline Brixey^a, Juliana J. Brixey^b

^a Department of Computer Science, University of Southern California, Los Angeles, CA, USA,

^b School of Biomedical Informatics and School of Nursing,
The University of Texas Health Science Center at Houston, Houston, TX, USA

Abstract

The Rede NUTES telemedicine question submission system provides second opinions to remote healthcare professionals in the northeastern state of Pernambuco, Brazil. Submitted questions to the telemedicine system by general practitioners and nurses were analyzed using big data exploration techniques to summarize topic, trends and lexical features.

Keywords:

Telemedicine; Brazil, Mining

Introduction

The Brazilian constitution guarantees free public healthcare to every citizen, however, many specialists are concentrated in the large coastal cities of the country, resulting in an underserved rural population [1]. The Health Ministry of Brazil implemented the Programa Telessaúde Brasil Redes to combat this issue by providing health support, advice from specialists, and permanent health education via telemedicine. Telemedicine can connect any member of the health care team across the continuum of care to enhance outcomes. For example, the Rede offers second opinions to submitted questions via their online submission system to be reviewed by other health care professionals [2]. The opportunity to seek second opinions allows for non-specialists to provide specialized service in remote areas, thus saving patients from undertaking a costly trip for care.

Data from the Rede NUTES submitted questions were analyzed and explored using big data methods to gain insights into how and for what Brazilian nurses and general practitioners (GPs) utilized the system.

Methods

The data set analyzed comes from text questions submitted from the years 2010-2012. In total, 5580 questions were submitted and answered. In the pre-processing phase, all identifying information was removed by hand to comply with HIPPA standards. The data exploration phase followed a four step process.

Step 1 was to identify and isolate questions submitted by nurses and GPs to compare how these groups used the Rede system in comparison to their peers. In step 2, the questions were reviewed by calendar months. This visualized usage trends over the course of a year. Step 3 was to determine question themes by month. By visualizing question topics and clustering those topics in a timeline, it was possible to

understand topic trends in the data. In the final step, the data were reviewed for lexical relationships using Linguistic Inquiry and Word Count (LIWC) program in Portuguese [3]. LIWC returns values on dimensions, for example the number of self-references and level of positivity, that insights about the state of mind of the writer.

Results

Nurses and nurse technicians submitted the most questions to NUTES for two out of the three years reviewed (38% in 2010 and 32% in 2011), and in all three years combined (36%). This group submitted 92% of all questions in December 2012. The top question topics submitted by nurses concerned cardiology, gynecology, obstetrics, pediatrics, and dentistry.

GPs did not use the question system as prolifically, and submitted only 6% of the total number of questions during the three years. Questions submitted by GPs largely focused on cardiology, such as queries about hypertension management.

Conclusions

Exploratory data analysis is the first step in using big data for future experiments. It also allows researchers to detect mistakes in the data that could potentially skew machine learning results, and check intuitions about the data. It is an essential step in telemedicine systems to evaluate and analyze trends for a topic and usage in order to better serve future users of these systems.

References

- [1] L.H. Iwaya, M.A. Gomes, M.A. Simplicio, T.C.M.B. Carvalho, C.K. Dominicin, R.R. Skauragui, ... & P. Hakansson, Mobile health in emerging countries: A survey of research initiatives in Brazil, *Int'l J Med Inform* **82**(2013), 283-298.
- [2] L.M.P. Sanches, Telessaúde: sistema de busca de casos clínicos para apoio à Estratégia Saúde da Família, *Doctoral dissertation, Universidade Estadual de Campinas. Faculdade de Ciências Médicas* (2013).
- [3] J.W. Pennebaker, M.E. Francis, R.J. Booth, Linguistic inquiry and word count: Liwc 2001, *Mahway: Lawrence Erlbaum Associates* (2001).

Address for correspondence

Jacqueline Brixey, brixey@usc.edu

Integrating x4T-EDC into an Image-Portal to Establish an Ophthalmic Reading Center

Philipp Bruland^a, Ulrich Kathöfer^b, Maximilian Treder^c, Nicole Eter^c, Martin Dugas^a

^aInstitute of Medical Informatics, University Hospital Münster, Münster, Germany,

^bITZ FuL, University of Münster, Münster, Germany,

^cDepartment of Ophthalmology, University Hospital Münster, Münster, Germany

Abstract

Reading centers provide centralized high-quality diagnostics in ophthalmic clinical trials. Since ophthalmic images are captured in electronic format at peripheral clinics, an integrated workflow for image transfer and creation of structured reports is needed, including quality assurance. The image portal and the study database are separate components. We assessed whether this integration is feasible with trial-related IT standards and built a prototype system as a proof-of-concept. CDISC ODM and OAuth authentication were used to integrate the image portal with x4T-EDC, facilitating automatic data transfer and single sign-on.

Keywords:

Electronic Data Capture, Reading Center, System Integration

Introduction

Assessment of ophthalmic images requires a standardized procedure, especially in clinical trials. Therefore, so called reading centers are established to receive images from participating study sites, assess those by medical experts in a standardized manner and document results in the study database [1]. Reading centers usually consist of an image portal into which images of study subjects are uploaded. Separate from the portal, a database is used for study documentation. In this regard, validated electronic data capture (EDC) systems are applicable that fulfil certain regulatory requirements for use in clinical trials. Nevertheless, such systems are usually separated IT components resulting in error-prone and time-consuming disruptions of the reading workflow.

Hence, the aim of this research is to investigate, whether it is feasible to connect an image portal with an EDC system using trial related IT standards.

Methods

System requirements were identified through meetings with clinical partners from the clinic of Ophthalmology. For this proof-of-concept the capabilities of an image portal and the x4T-EDC system [2] were analyzed. x4T-EDC is an in-house developed EDC system supporting mainly investigator initiated trials and diverse register projects. Different formats for data transfer between the portal and the study database (x4T-EDC) were reviewed such as CSV files or the XML-based format ODM (Operational Data Model) of the CDISC (Clinical Data Interchange Standards Consortium) [3]. To prevent users entering their credentials into both systems, single sign-on mechanisms such as OAuth were examined.

Results

ODMs "ClinicalData" part was used to transfer basic information of the uploaded image via web services, like visit number, date, eye location, image type and recording device from the portal to the x4T-EDC system. Single sign-on was established using OAuth-similar authentication mechanisms. The x4T-EDC system was enhanced towards a more specific right-management for preventing readers to access the reports of other readers. x4T-EDC compares the reports of both readers to support quality assurance performed by a senior reader. Possible deviations are presented by x4T-EDC to the senior reader and again examined. Information regarding image status is communicated between the portal and x4T-EDC to support the reading workflow.

Discussion

Defined interfaces allow the integration between the reading portal and the study database. Based on the well-established and widespread format CDISC ODM, which is used by most EDC solutions, the exchange of data is realized. Single sign-on mechanisms allow a slim and clear process from the examination to the assessment of images. The separation of both systems (image-portal accessible via internet, study-database only via intranet) is an advantage regarding security. At present, the implementation of the reading-center is a successful proof-of-concept. We also intend to perform an usability evaluation with the system usability scale.

Conclusion

Integrating x4T-EDC into an image-portal with trial-related IT standards and technologies is feasible.

References

- [1] S.M. Hudson and et al, Centralized Reading Center Improves Quality in a Real-World Setting. *Ophthalmic Surg Lasers Imaging Retina*. 2015 Jun;46(6):624-9.
- [2] P. Dziuballe and et al, The single source architecture x4T to connect medical documentation and clinical research. *Stud Health Technol Inform*. 2011;169:902-6.
- [3] Operational Data Model. <http://www.cdisc.org/odm>. Accessed: November 2016

Address for correspondence

Philipp Bruland: philipp.bruland[at]uni-muenster.de

Technical Challenges and Opportunities When Implementing Pharmacogenomics Decision Support Integrated in the Electronic Health Record

Pedro J. Caraballo^a, Joseph A. Sutton^b, Ann M. Moyer^c, David Blair^d, Lois C. Hines^e, Padma S. Rao^b, Mark F. Adams^b, Sahana Murthy^b, Tina Garza^b, Mary E. Karow^e, Harmanjit Singh^b, Jyothsna Giri^f, Donald B. Gabrielson^g, Jennifer St. Sauver^h, Suzette J. Bielinski^h, Mark A. Parkulo^a

^a Department of Medicine

^b Department of Information Technology

^c Department of Laboratory Medicine & Pathology

^d Department of Family Medicine

^e Department of Pharmacy Services

^f Center for Individualized Medicine.

^g Department of Management Engineering

^h Department of Health Sciences Research Mayo Clinic, USA

Abstract

Clinical use of pharmacogenomic (PGx) knowledge at the bedside is new and complex. Our program has implemented multiple PGx-CDS interventions in different clinical settings and in multiple commercial EHRs. Herein, we discuss lessons learned and propose general technical guidelines related to PGx implementation.

Keywords:

pharmacogenetics, clinical decision support systems, electronic health records

Introduction

Precision medicine will revolutionize the current model of healthcare to allow the delivery of personalized care based on multiple predictors including individual genetic information. Pharmacogenomics (PGx), the study of the role of genetics in drug response, is one aspect of precision medicine that promises a significant impact to the general population. In fact, the vast majority of the population has at least one variant in genes associated to the metabolism or mechanism of action of drugs (pharmacokinetic and pharmacodynamics) or immune-related genes that increase risk of hypersensitivity reactions. Although many challenges have hindered the implementation of PGx into clinical practice, it is now increasingly evident that the time has come for widespread adoption [1].

PGx-CDS Implementation

The Mayo Clinic Center for Individualized Medicine has implemented a Pharmacogenomic Program to translate genomic knowledge into clinical practice. Using institutional resources and a comprehensive operational model [2, 3], a multidisciplinary group of experts developed and implemented multiple CDS interventions integrated in two EHRs to alert for 19 drug-gene interactions.

Lessons Learned

During the PGx-CDS implementation, many challenges and opportunities were identified. We have grouped these in several main themes.

1. Integrated multidisciplinary team approach: PGx-CDS implementation should not be a standalone project.

2. Structured format of PGx test results: Translation of the PGx test reports into structured format (discrete results or constrain text) in the EHR is critical for implementation.
3. Complementing EHR functionalities: Current EHR functionality is sufficient, though not perfect to implement PGx-CDS. Pop-up alerts should not be the only intervention.
4. Clinical champions and education: Clinical champions from different specialty areas should be invited to participate in the development of PGx-CDS to help decrease clinical resistance
5. Establish a process for ongoing maintenance: Long-term maintenance cannot be underestimated. PGx-CDS needs very frequent updates due to changes in laboratory methodologies, discovery of novel variants, new medications, and emerging research evidence resulting in changes to the interpretations of existing variants and PGx guidelines.

Conclusion

PGx-CDS implementation is feasible by using current commercially available EHR systems. However, PGx is an evolving science and the PGx-CDS interventions require constant review and updates. PGx-CDS implementation benefits from a multidisciplinary infrastructure supporting not only the implementation team but also the clinicians receiving the CDS alerts.

Acknowledgements

The authors thank the Mayo Clinic Center for Individualize Medicine and the Office of Information and Knowledge Management for their support. The authors report no conflict of interest.

References

- [1] M.V. Relling, W.E. Evans, Pharmacogenomics in the clinic, *Nature* **526** (2015), 343-350.
- [2] S.J. Bielinski, J.E. Olson, J. Pathak, et al., Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time-using genomic data to individualize treatment protocol, *Mayo Clin Proc* **89** (2014), 25-33.
- [3] P.J. Caraballo, L.S. Hodge, S.J. Bielinski, et al., Multidisciplinary model to implement pharmacogenomics at the point of care, *Genet Med* (2016).

Address for correspondence

Pedro J. Caraballo, MD. Mayo Clinic. 200 First Street SW. Rochester MN, 55905. USA. Phone: 507-2848586. caraballo.pedro@mayo.edu.

Validated Simulation: The Preliminary Experience of Anesthesiologist Board Examination in Taiwan

Hung-Wei Cheng^{a,b,d}, Pin-Tarng Chen^{a,b,*}, Ying-Chih Lo^{c,d}, Po-Lun Chang^d

^a Department of Anesthesiology, Taipei Veterans General Hospital, Taipei, Taiwan, China

^b Department of Anesthesiology, National Yang-Ming University School of Medicine, Taipei, Taiwan, China

^c Division of Nephrology, Department of Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, China

^d Institute of Biomedical Informatics, National Yang-Ming University, Taipei, Taiwan, China

*Co-first author, contributed equally to this work

Abstract

High fidelity simulation-based teaching has played an important role in medical education, especially in anesthesiology and emergency. But there is not any currently validated scoring system or prediction model for high fidelity simulation. We will develop a validated prediction model to enhance the efficiency and validation of clinical training with high fidelity simulation.

Keywords:

Anesthesia; Education; Medical

Introduction

High fidelity simulation-based teaching currently has played an important role in medical education, especially in anesthesiology and emergency. After several years of pilot exercise, Taiwan Society of Anesthesiologist have formal put high fidelity simulation into part of board examination. After literature review, there is not any currently validated scoring system or prediction model for high fidelity simulation. Therefore, development of a validated scoring or model is an important task for us to keep teaching with high fidelity simulation.

Methods

We had collected the data of anesthesiologist board examination in 2013. De-indentification was also done at the same time for data privacy. The typical simulation scenario we used to build relational database is perioperative myocardial infarction (MI). It's a rare but fatal complication during operation. Early vigilance and appropriate treatment is the key to save patient's life.

50 chief residents of anesthesiology had took this examination. Their performance were rated with a check list of 41 technique-related items and 8 non-technique items by 2 independent rater. Final outcome of each examinee is a global rating score from 1 to 10.

Eighty percent of the dataset were used as training data, and remaining were used for model testing. We utilized various machine learning algorithms, such as decision tree (DT) with the implement of Categorical and Regression Tree (CART), random forest (RF), support vector machine (SVM) with different kernel and linear regression. The model's performance was evaluated with root mean square error and Pearson correlation coefficient. Feature selection was also done with random forest.

Results

All machine learning algorithms that we used have good performance in prediction, but SVM with linear kernel and

random forest are slight better than others with correlation coefficient 0.99 and 0.98, respectively. Top 7 important features ranked by random forest include 3 technique features and 4 non-technique features as table 2.

Table 1. Performance of various machine learning algorithms on testing data

Algorithm	RMSE	Correlation coefficient
Decision Tree (CART)	0.65	0.86
Random Forest	0.24	0.98
SVM (Linear kernel)	0.18	0.99
SVM (Polynomial kernel)	1.20	0.92
SVM (RBF kernel)	0.62	0.94
Linear regression with feature selection (Top 7)	0.28	0.98

CART: classification and regression tree, SVM: support vector machine, RMSE: root mean square error

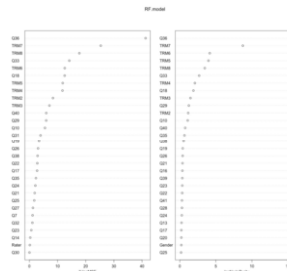


Table 2.

Feature code	Meaning
Q36	Correct diagnosis of MI
TRM7	Re-evaluation after treatment
TRM8	Discuss treatment with surgeon
Q33	Order 12-leads ECG
TRM6	Follow up orders
Q18	Set correct joules in defibrillation
TRM5	Order drug, dosage, and route clearly

Conclusions

In this preliminary study, we find both SVM with linear kernel and RF could build accurate prediction model in this simulation test. Otherwise, according to the result of feature selection by RF, evaluation of non-technique factors may play a more important role than before. Due to limited case number, larger study maybe conducted to confirm the prediction accuracy in the future.

References

- [1] D.M. Gaba, The future vision of simulation in health care. *Qual Saf Health Care* 13(2004), i2-i10.
- [2] H. Berkenstadt, A. Ziv, N. Gafni and et al, Incorporating simulation- based objective structured clinical examination into the Israeli national board examination in anesthesiology. *Anesth Analg* 102(2006), 853-858.

Address for correspondence:

Polun Chang, Institute of Biomedical Informatics, National Yang Ming University, No.155, Sec.2, Linong Street, Taipei, 112 Taiwan; polun@ym.edu.tw

How to Use TCM Informatics to Study Traditional Chinese Medicine in Big Data Age

Shi Cheng, Gong Qing-yue, Zhou Jinhai

College of Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

Abstract

This paper introduces the characteristics and complexity of traditional Chinese medicine (TCM) data, considers that modern big data processing technology has brought new opportunities for the research of TCM, and gives some ideas and methods to apply big data technology in TCM.

Keywords

Electronic Health Record; Research; Medicine Chinese, Traditional.

Introduction

Chinese medicine believes that disease is the result of the struggle between “righteousness” and “evil”. All that are conducive to the body against disease are considered “righteousness”, things not conducive to human health survival are “evil”. Classification of disease contains human information and insight into people’s understanding of disease in the macro environment. In the course of thousands of years, Chinese medicine has accumulated rich experience and knowledge in the application of acupuncture, massage, and other methods of diagnosis and treatment. These big data are an extremely valuable asset worth exploring. Thousands of traditional Chinese medicine hospitals in China have adopted an Electronic Medical Record system to record clinical data. Accumulating over a long period, Chinese medicine data is amazing, containing conditions worth further exploring in the era of big data technology.

Methods

Chinese medicine is considered as “unscientific” due to the lack of objective data and standardized scientific language. These shortcomings make traditional Chinese medicine information processing difficultly.

Traditional Chinese medicine uses the doctor’s sensory organs to collect patient data (subjective data of doctors) and most of the data recorded by inquiry describes the patient’s subjective

sensory. This data is recorded in natural language. Traditional computer data is objective. Modern big data technology (e.g., Google Flu) provides a new opportunity to use computer technology to study subjective traditional Chinese medicine data.

Results

We believe that in a pilot implementation, we can choose several Chinese medicine hospitals and suitable clinical research projects for early exploration. Traditional Chinese Medicine clinical data can be standardized in accordance with the requirements of the State Administration of Traditional Chinese Medicine. This pilot will allow us to accumulate experience and methods about modernization of Chinese medicine.

Conclusion

The big data age provides new insight that could not be acquired with smaller data sets. Modern large-scale data processing technology has brought new opportunities for traditional Chinese medicine research. In this paper, the characteristics and problems of traditional Chinese medicine data were analyzed and new ideas and methods were proposed.

References

- [1] C. Shi, Present situation and characteristics of traditional Chinese medical information, In: *The Ninth International Annual Conference on Medical Information*, Seoul Korea (1988):841–84.
- [2] C. Shi, L.S. Xu, Q.Y. Gong, X.D. Wang, The characteristics and key issues in electronic medical records (EMR) of traditional, MEDINFO 2013 (2013):976
- [3] C. Shi, Q.Y. Gong, J.H. Zhou Traditional Chinese medicine (TCM) and TCM information, In: *World Chinese Health Informatics Symposium* (2015):28–29

Address for correspondence

Shi Cheng: cshi2005@126.com; Gong Qing-yue: qygong@126.com; Zhou Jin-hai: zhoujh2003@126.com

Identifying Use Cases for Electronic Nursing Record Systems Using Clinical Workflow Observations and a Delphi Survey

Insook Cho^a and Joohyun Chung^b

^a Department of Nursing, Inha University, Incheon, Republic of Korea,

^b College of Nursing, University of Massachusetts, Dartmouth, USA

Abstract

This study describes the process of identifying use cases for next generation electronic nursing records in a high-level view. Literature review, clinical observation, and a Delphi survey were employed. Eight use cases were identified with importance and convergence scores.

Keywords:

Nursing records; Medical records systems, computerized.

Introduction

Electronic nursing record (ENR) systems have been adopted world-wide as an essential part of clinical practice. Due to the lack of a knowledgeable nursing executive and clinician workforce regarding standardized processes and data, most ENR systems are not based on explicit and implicit models for information and content standards. Instead, they are based on paper-view models. This has hidden many of the benefits of an ENR, such as supporting efficient, complete, and accurate information in practice, support of quality improvement, analytics, and research [1]. In addition, nurses' dissatisfaction with ENRs has been frequently reported as those are not aligned with workflow; it is critical that clinical work process be supported. We aimed to understand and delineate a clinical work process from the end users' point of view.

Methods

A Literature review of nursing documentation standards and guidelines, EHR usability, and ISO S/W quality specifications was performed. Clinical observations took place at 24 medical and surgical units from six tertiary hospitals located in a metropolitan area of South Korea. A time-motion method was used and 72 nurses from 24 units at each hospital, plus an additional 32 non-participating nurse observers were included. Out of 108 sessions, we randomly selected 21 (20%) samples for qualitative analysis. The patterns were described and abstracted to several use cases. Each use case was structured by definition and requirements on functional, data and information, and design. In a Delphi survey, two panels were recruited: (1) six nurse managers from the participated hospitals, plus two nursing directors; (2) three academic nursing informaticians (NI).

Results

Figure 1 shows the eight high-level use cases identified from clinical observation pattern analysis. In Delphi survey, the *Daily care plan & set goals* and the *Evaluation* use cases showed lower importance and convergence scores.

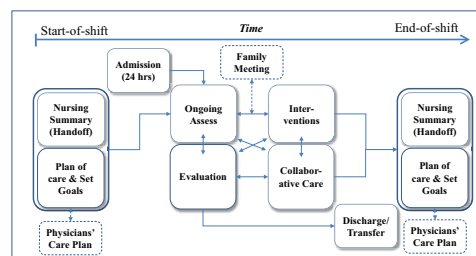


Figure 1 - High-level Use Cases

Table 1 - Importance and convergence scores by use case

Use Cases	Importance (0-5)	Convergence (%)
Nursing summary (Handoff)	5	70%+
Plan of care & set goals	4	60%
Admission (24 hours)	5	70%+
Ongoing assessment	5	70%+
Intervention	5	50%
Collaborative care (physician order)	5	70%+
Evaluation (reassessment)	4	50%
Discharge/Transfer	5	60%

Discussion

The *Plan of care & set goals* and *Evaluation* use cases were less important. There were many different perspectives for defining the intervention use case specifications, even when considering importance scores. We also found gaps between nurse managers, NI experts, and nurse managers.

Conclusion

We identified eight important use cases for clinical practice.

Acknowledgements

This study was supported by a grant of the Korea Healthcare Technology R&D Project, Ministry for Health, Welfare & Family Affairs, Republic of Korea (No. HI15C1089).

References

- [1] B.L. Westra, G.E. Latimer, S.A. Matney, J.I. Park, J. Sensmeier, R.L. Simpson, M.J. Swanson, J.J. Warren, C.W. Delaney, A national action plan for sharable and comparable nursing data to support practice and translational research for transforming health care. *J Am Med Inform Assoc* 22 (2015), 600-607.

Address for correspondence

Insook Cho, PhD. 100 Inha-ro, Namgu, Incheon, Republic of Korea, 22212.

Delivering Digital Drugs: An Exploratory Study of the Digitalisation of Supply and Use of Medicines

Tony Cornford^a, Valentina Lichtner^b, Jane Dickson^a, Ralph Hibberd^a,
Ela Klecun^a, Will Venters^a, Bryony Dean Franklin^c

^a Department of Management, London School of Economics, London, UK

^b University of Leeds Business School, University of Leeds, Leeds, UK

^c UCL School of Pharmacy, University College London, London, UK

Abstract

Medicines' supply and use is increasingly reliant on digital means and information. This poster presents exploratory research over five episodes of digitalisation of medicines across the supply network. We 'follow the drug' through this emerging field, providing an initial map of this new territory.

Keywords:

Drug utilization; Information systems; Precision medicine.

Introduction

It was all very well to say 'Drink me,' but the wise little Alice was not going to do that in a hurry. 'No, I'll look first,' she said, 'and see whether it's marked "poison" or not' (Alice in Wonderland, Chapter 1)

The Delivering Digital Drugs (D3) project is concerned with the accelerating digitisation of the chain of production and consumption of therapeutic drugs (medicines) and the emergence of what we term 'digital drugs' [2].

Health care is increasingly reliant on medicines as a primary means of therapy and medicines are increasingly developed and delivered by reliance on associated and accompanying data and information technologies within a rich and rapidly changing data ecosystem. Today a medicine can be seen as a hybrid artefact, part chemical, part data, part supply chain and logistics, part knowledgeable practice that is itself part embodied and part algorithmic. This raises many complex intersecting issues including: what drives this process of digitalisation; what constitute appropriate functional uses of data - e.g. in prescribing, in research, in audit, in monitoring individual or population wide effectiveness or value for money, in monitoring side effects etc. There are also substantial new issues of data stewardship, openness and ethical data governance practices. At the core are the questions of what is or may become deemed appropriate professional practice, taking into account issues of safety and avoidance of errors, as well as the status and role of patients as active participants in medicines use.

Methods

Our methodology is cross-disciplinary and exploratory. It includes five case studies of the processes of digitalisation sampled along the chain of supply and use of medicines:

1. Anti-counterfeiting in the medicines supply networks,

2. Hospital prescribing and supply of drugs and related data infrastructures,
3. Pharmacovigilance and the monitoring for adverse effects of therapeutic drug use,
4. Research data services, principally the aggregation of patient data for secondary use,
5. Patient views' of use of medicines and of technology to support patients in their role as active participants in their therapy.

A mix of qualitative data collection methods were applied: interviews, observations and documentary analysis. The cases are explored both as single episodes and in their interconnections. In order to facilitate the latter, we apply a tracer approach [1] and 'follow the drug' through and between the cases. Drugs are taken as 'tracers' that can reveal digitalisation paths and flows within and across cases. Thus the 'digitalization of medicines' is studied 'following the drugs' from the moment they enter the market for clinical use, through their transit through pharmaceutical supply networks [case 1] to the process of prescribing and dispensing [case 2], to patient's decision making [case 5], to the capture of aggregated data around the use of therapeutic drugs [case 4] and the outcomes of this use [case 3-4].

Acknowledgements

Project funded by Research Councils UK / 'New Economic Models in the Digital Economy' Programme (EP/L021188/1).

References

- [1] P. Combe, A tracer approach to the study of organizations, *Journal of Management Studies* 17 (1980), 96-126.
- [2] T. Cornford and V. Lichtner, Digital Drugs: An anatomy of new medicines, in: *Information Systems and Global Assemblages: (Re)configuring Actors, Artefacts, Organizations. IFIP Advances in Information and Communication Technology.*, B. Doolin, E. Lamprou, N.N. Mitev, and L. McLeod, eds., Springer, Auckland, New Zealand, 2014, 149-162.

Address for correspondence

Tony Cornford, t.cornford@lse.ac.uk

Concept Embedding for Relevance Detection of Search Queries Regarding CHOP

Yihan Deng^a, Lukas Faulstich^b, Kerstin Denecke^a

^a Bern University of Applied Science, Bern, Switzerland,

^b ID Information and Documentation GmbH, Berlin, Germany

Abstract

Automatic encoding of diagnosis and procedures can increase the interoperability and efficacy of the clinical cooperation. The concept, rule-based and machine learning classification methods for automatic code generation can easily reach their limit due to the handcrafted rules and a limited coverage of the vocabulary in a concept library. As the first step to apply deep learning methods in automatic encoding in the clinical domain, a suitable semantic representation should be generated. In this work, we will focus on the embedding mechanism and dimensional reduction method for text representation, which mitigate the sparseness of the data input in the clinical domain. Different methods such as word embedding and random projection will be evaluated based on logs of query-document matching.

Keywords:

Automatic Encoding, Classification, Machine Learning

Introduction

In order to claim costs to the health insurance and for clinical documentation purposes, it is necessary and even legally required to encode diagnoses and procedures by classification codes from relevant classification systems. In Switzerland, these are ICD-10-GM for diagnoses and Schweizerische Operationsklassifikation (CHOP) for medical procedures. In order to facilitate the subsequent query matching process based on deep learning, we investigate the possibilities of direct embedding of concept through word2vec principle (Skip-gram, continuous bag of words, negative sampling) as well as the dimension reduction with the input of sparse concept vector. We would like to figure out whether the embedding method can be applied on the concept vector directly and which type of embedding (Skip-gram and CBOW) is more suitable for the concept embedding from domain specific data.

Methods

For test and development, we are using semantic representations in vector form generated from search entries and, on the other hand, target catalog texts that have been generated from the CHOP classification texts. As is illustrated in Figure 1, for each query – classification code text pair it has been assigned whether the classification text matched the user query or not. For generating the vectors, search entries and classification texts have been mapped to concepts of a medical terminology by the terminology server ID MACS®. As pre-processing for the query matching system, the input layer is in charge of the representation generation and dimension reduction. This layer is trying to represent the word in a corpus by a special instantiation of a set of hidden variables. The

embedding process learns the representation of each word by maximizing the log likelihood of each word given its context (context window). Our evaluation platform will test the embedding result based on CBOW, Skip-gram, and random projection. The embedding is implemented with Tensorflow [1].

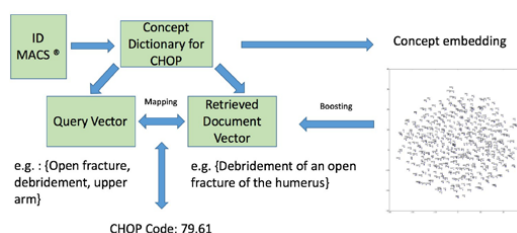


Figure 1- Concept embedding for a better automatic CHOP encoding.

Results

20067 logs are used to do the embedding and evaluation. Based on the annotated relevance metrics [1], the average match rate of Skip-gram embedding based on concept only vector representation has achieved the best relevance match (0.63), while the CBOW has achieved clearly less match rate (0.43). The random project has only reached the least match rate around 10%.

Discussion and Conclusion

Concept embedding can largely reduce the sparseness to make a suitable input for the deep neural network. The skip-gram is most suitable method for encoding short text queries referring to clinical or surgical procedure, since keywords are short and relatively independent. The portability and scalability of the method will be tested in other medical natural language text.

Acknowledgements

This work is supported by ID Information and Documentation GmbH, Germany

References

- [1] Schnabel T, Labutov I, Mimno D, Joachims T. Evaluation methods for unsupervised word embeddings 2015, EMNLP, 298-307.

Correspondence:

Yihan Deng
yihan.deng@bfh.ch

¹ <https://www.tensorflow.org>

Classifying Clinical Notes with Pain Assessment

Samah J. Fodeh^a, Dezon Finch^b, Lina Bouayad^b, Stephen Luther^b, Robert D. Kerns^a, Cynthia Brandt^a

^aYale Center for Medical Informatics, Yale University School of Medicine, New Haven, CT, USA

^bHSR&D Center of Innovation on Disability and Rehabilitation Research, James A. Haley Veterans Hospital, Tampa, FL, USA

Abstract

Pain is a significant public health problem, affecting an estimated 100 million Americans. Evidence has highlighted that patients with chronic pain often suffer from deficits in pain care quality (PCQ). Efforts to improve PCQ hinge on the identification of reliable PCQ indicators such as pain assessment. In this study, we developed a classifier that leverages narratives in clinical notes to derive indicators of pain assessment for patients with chronic pain.

Keywords:

Pain, classification.

Introduction

Currently, there is no intelligent and reliable approach to identify patients with clinically significant pain in the EHR [1]. Capturing data elements related to pain in most EHRs is not standardized. While a growing number of pain-related studies have utilized structured and easily retrievable coded fields in the EHR, none have explored the utility of provider narratives in clinical notes. We employed machine learning (ML) algorithms to analyze unstructured narrative text data in the EHR to develop a reliable classifier that identifies clinical notes with pain assessment

Methods

Using data from the Veterans Health Administration (VHA) system, we abstracted a random sample of 1058 primary care notes from the medical records of patients with initial Musculoskeletal Diagnoses (MSD) in 2011. The views expressed in this poster are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs. We manually annotated the data and generated a reference standard for the purposes of building a classifier that labels clinical notes with indicators of pain assessment as positive and other notes as negative. Pain assessment information that was annotated in the notes include terms related to intensity, quality, persistence, diurnal variation, aggravating factors, alleviating factors and functional assessment. Inter annotator agreement was 72%. A note that included any of the relevant terms was deemed positive. We experimented with multiple classifiers including decision tree (DT), support vector machine (SVM), k-nearest neighbor (KNN), and random forest (RF). We used two-thirds of the data for training and one-third for testing. We performed 10-fold cross validation to train the classifiers and measured performance via accuracy, positive predictive value, sensitivity, F1-score and area under the curve (AUC).

Results

Table summarizes the average performance of the different classifiers on the training and test sets in all measures. On the training data, the RF classifier did best in most measures; the highest F1-score and area under the curve were achieved .94 and .94, respectively. Similar performance was observed on the test set, nonetheless, the K-nearest neighbor classifier outperformed RF in terms of sensitivity. We used 6 neighbors to estimate the label of a given clinical note (i.e. k=6); adding beyond that did not seem to change the results. KNN, however, still did worse than all classifiers in terms of the other remaining measures for both training and test data sets as shown in the table. The best performance of RF is explained by the combined predictions of several base classifiers which likely improved generalizability and robustness over a single classifier. In our experiments we combined 100 learners for prediction.

Training Data					
	Accuracy	PPV	Sensitivity	F1-score	AUC
SVM	0.88	0.87	0.92	0.89	0.88
RF	0.94	0.96	0.93	0.94	0.94
KNN	0.85	0.80	0.95	0.87	0.87
DT	0.92	0.93	0.92	0.93	0.92
Test Data					
SVM	0.88	0.82	0.97	0.89	0.89
RF	0.94	0.95	0.94	0.94	0.94
KNN	0.85	0.78	0.99	0.87	0.88
DT	0.91	0.89	0.94	0.91	0.91

Conclusion

The developed pain assessment classifier is potentially useful for improving quality of pain care. We vision it as an agent that can automatically sift through the EHR to pull clinical notes with pain assessment for further pain quality research. The RF classifier outperformed other single classifiers. In future work, we will build a more granular classifier that determines what type of assessment has been done on patients in clinical notes.

Acknowledgement

This study was funded by NIH – grant number (1R01AT008448-01).

References

- [1] T.Y. Tian, Zlateva I., and Anderson D.R., *Using electronic health records data to identify patients with chronic pain in a primary care setting*. Journal of Medical Informatics, 2013. **20**(e2): p. e275-e280.

Address for correspondence:

Samah Fodeh (Samah.fodeh@yale.edu)

Modeling an IT Support for Handling Serious Adverse Events in Clinical Trials

Simon Fonck, Thomas Deserno

Department of Medical Informatics, Uniklinik RWTH Aachen, Aachen, Germany

Abstract

Serious adverse events (AE) or reactions (AR) may occur in clinical trials and require particularly regulated reporting. Manual management is inefficient and ineffective. Based on a description of regulations, we have developed a data model with class-, state-, use-case- and activity diagrams, which can be used for automatic code generation of an assisting software tool for AE / AR data management.

Keywords:

Clinical Trial, Adverse Event, Software Engineering

Introduction

Several IT systems are used to support clinical trial conduction but they are usually disconnected from the hospital information systems (HIS). Therefore, the principle investigator of a controlled clinical trial might not be informed when a serious adverse event (SAE) or a suspected unexpected serious adverse reaction (SUSAR) occur. To ensure information exchange between the HIS and the clinical trial management system (CTMS), we model an appropriate IT infrastructure supporting SAE / SUSAR identification. Furthermore, the tool is designed to support and assist the trial team in timely and accurate reporting of such events.

Methods

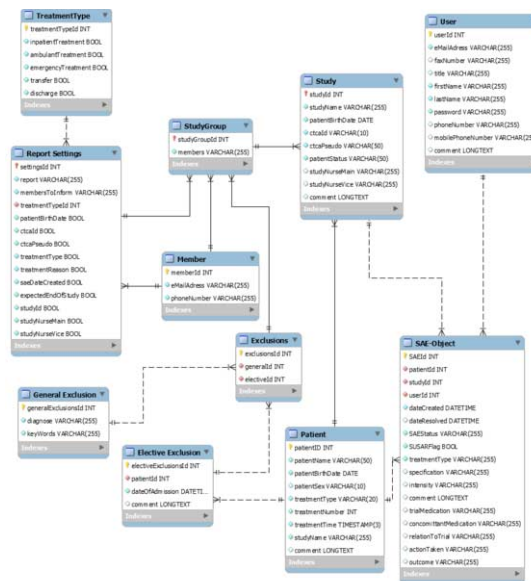
There are several software engineering tools, such as the unified modeling language (UML) [1] used to describe the functional requirements for the SAE tool and the occurring processes within models provided by the UML. Based on a comprehensive description of regulatory requirements and a desired workflow [2], we created

- Use-case diagrams to describe functionalities from the user's view;
- Activity diagrams to represent the data- and controlflow;
- Class diagrams to define the structure of the system;
- State diagrams to compose a finite number of system states.

These models serve as basis for the automated code generation [3] of an web-based SAE management tool, its documentation, and the verification of the implemented source code.

Results

The SAE management is divided into two processes: automatic SAE identification and manual monitoring of SAEs including parallel automatic support. In a relational data model, only eleven tables are required.



Discussion

Comprehensive modeling supports the developer during the implementation process. The software will simplify the management of SAEs in clinical trials. Automatically generated messages guide research nurses to keep track of SAEs, lower the risk of missing a case, and increase security and quality of clinical trials in general.

References

- [1] Rumpe B. Modeling with UML: Language, Concepts, Methods. Springer International, 2016
- [2] Pelzer A, Wnendt S, Deserno T, Deserno V. Konzeptentwicklung zur IT-gestützten Erfassung, Dokumentation und Meldung von schwerwiegenden unerwünschten Ereignissen in eigeninitiierten Arzneimittelstudien. GMS Publishing House 2015; 60: 052. [german]
- [3] Monticore DEx Software Engineering. URL <http://www.monticore.de/dex/>, accessed: 12.12.16.

Addresses for correspondence

simon.fonck@rwth-aachen.de; deserno@ieee.org.

A Study on Data-Driven Novel Cancer Staging Methods

Yuan Gao, Yu Tian, Shengqiang Chi, Yao Lu, Xinhang Li, Tianshu Zhou, Jing-song Li

Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou Zhejiang Province, China

Abstract

This paper presents a data-driven method to study the relationship of survival and clinical information of patients. The machine learning models were established to study the survival situation at the time of interest based on survival analysis. The way to determine the time of interest is an innovation of this paper. The distribution of survival time is considered, namely the three quartiles, as well as the traditional analysis experience is taken into consideration.

Keywords:

Colorectal Neoplasms, Machine Learning, Survival Analysis

Introduction

According to the recent studies, the widely used 7th AJCC TNM staging system cannot make a linear prediction on survival, which means the survival rate is not getting smaller as the staging gets higher, but there is a fluctuation[1]. This paper attempted to implement data-driven machine learning algorithms, built several different models, and finally, the resulting models were assessed and compared. This paper concluded three quartiles of the survival rate of the patients from data above using survival analysis[2], which are 3, 5, and 7 years after diagnosis. The idea of research of this paper is not from experience but from the data itself, which provides a new way of thinking of research on cancer staging method.

Methods

The data used in this paper is the follow-up data from the 2080 patients admitted in 1985-2011 by the Department of Surgical Oncology, Second Affiliated Hospital, Zhejiang University College of Medicine. Multivariate Imputation by Chained Equations based on random forest model is introduced[3] to process the missing values. Logarithmic transformation is used to remove the effect of outliers and Min-Max Normalization is used to balance the contribution of different features. Through the survival analysis of the original data, the overall distribution of the data is obtained: three quartiles of the survival time, which are the time of interest points we need. In this paper, three machine learning models suitable for binary classification are selected: logistic regression model, support vector machine model and random forest model.

Results

According to the result of survival analysis, 3 quartiles of survival time was obtained. They are 40 month, 59 month, and

85 month. This paper identified the time points of interest are 3 year (36 month), 5 year (60 month) and 7 year (84 month) in order to take into account the general experience of the time points of study. According to the accuracy, the prediction of the models of 3-year survival performs the best overall, followed by that of 7-year and 5-year. It is the SVM model of the radial kernel function (3-year), and the logistic regression model (5-year and 7-year) that used to show the relatively good performance at each time point. However, if we take into account the specific circumstances of the prediction, the accuracy of negative samples at 3 years is much higher than that of positive samples. Meanwhile, the time of 7 years is the opposite, and the time of 5 years is more balanced.

Conclusions

Although the final result is not ideal, this paper has guidance for the future work, because it provides a complete data analysis process relatively. In the future work, with the acquisition of more clinical information and genomic information, it is believed that the result of this method could be improved.

Acknowledgements

This work was supported by Chinese National High-tech R&D Program (No.2015AA020109), National Key Scientific Instrument and Equipment Development Project (No.2016YFF0103200), and the Fundamental Research Funds for the Central Universities of China.

References

- [1] Jun Li Bao-Cai Guo Li-Rong Sun Jian-Wei Wang Xian-Hua Fu Su-Zhan Zhang Graeme Poston Ke-Feng, TNM staging of colorectal cancer should be reconsidered by T stage weighting, *World Journal of Gastroenterology* **20** (2014), 5104-5112.
- [2] R. Flynn, Survival analysis, *Journal of Clinical Nursing* **21** (2012), 2789-2797.
- [3] P. Li, E.A. Stuart, and D.B. Allison, Multiple Imputation: A Flexible Tool for Handling Missing Data, *JAMA* **314** (2015), 1966-1967.

Address for correspondence

Corresponding author: Jing-song Li. Address: College of Biomedical Engineering and Instrument Science, Zhejiang University, 38 Zheda road, Hangzhou 310027, China. Email: lis@zju.edu.cn. Tel.: +86-571-87951564

Methods for Faster and Efficient Data Entry in Electronic Medical Records

Shashi B. Gogia^a, Anand N. Malaviya^b

^a Software Developer and Surgeon In-charge, Sanwari Bai Surgical Centre, New Delhi, India

^b Consultant Rheumatologist, Indian Spinal Injuries Centre, New Delhi, India

Abstract

Data entry remains the slowest link in the value chain inhibiting growth of Electronic Medical Records and attendant benefits towards Meaningful use. We designed templates for user forms customized by specialty. Here, we demonstrate the functionality of our software and provide instructions on how these can be adopted by other developers.

Keywords:

Meaningful Use; Electronic Health Records; Software.

Introduction

While Moore's law assisted by creating memory that allows faster speeds for Information Systems, manual data entry methods remain relatively static. This is the slowest link in the value chain and is inhibiting growth of Electronic Medical Records (EMRs). Short cuts like Copy Paste are frowned upon. Issues important for IT personnel may not be viewed the same by clinicians (e.g., security protocols may not be essential and bog down the user).

The authors are clinicians developing software for more than 20 years. We have templates for user forms which can be customized quickly as per specialty. The software has list selects customized by the user. (e.g., Rheumatoid Arthritis (M05/6.x) ICD code specification triggers the DAS 28 (Disease Activity Scoring for 28 joints) screen with a homunculus image directing the user to the joint of interest). See Figure 1.

Blocks of expected repeated information become one-time entries by the clinician which can be edited. The software intelligently pastes text using keywords and permits editing at multiple layers depending on privileges. Some learning and assistance by the vendor or local IT team, is however required. Linkages to standardised codes can be done.

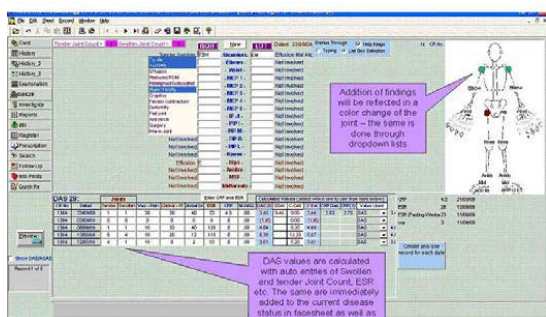


Figure 1 - A screen shot with multi-select drop-down lists and homunculus

Methods

We demonstrate the above and provide instruction for adoption

by developers. We have also analyzed user outcomes from Rheumatology. There are recordings of total patients and time spent in clinic before and after the software was adopted. Additionally, a survey was administered to patients comparing their overall experience during different stages of software development [1,2].

Results

With use of such techniques, total time spent by the patient in Rheumatology OPD decreased from 120 minutes to 45 minutes for the first visit and from 45 minutes to 25 minutes for revisits. The number of patients given appointment as well as seen, increased. Similar but less dramatic outcomes have been found with other specialties.

Conclusion

Adoption of EMRs and EHRs remains slow despite incentivization [3]. Our example results in better clinician engagement. Helpful features include:

- User created lists which allow auto completion and filtered selects via list boxes.
- Nicknames and codes which translate to bulk text.
- Multi-select list boxes with user created choices (e.g., nicknames).
- Bulk entry across linked columns (e.g., autofilled address information).
- Block standardized text entries (e.g., Operation notes).
- Labels and attributes are organized upon selection.

The system displays information understandable to the user. Besides ICD, physicians did not prefer other codes, even though available. Analysis can be performed with simple SQL. The scope can be configured for other specialties and diseases.

Users are the best developers. EMR adoption can be eased through deeper user engagement and decreased data entry time.

References

- [1] A.N. Malaviya, S.B. Gogia, Development, implementation and benefits of a rheumatology-specific electronic medical record application with automated display of outcome measures, *Int J Rheum Dis*, **13** (2010):347-360.
- [2] A. Kalyani, S.B. Gogia, A.N. Malaviya, S.B. Gogia, Data extraction for research purposes EMR vs text, *Indian J Med Informatics*, **8** (2014):9-12.
- [3] D.C. Classen, D.W. Bates, Finding the meaning in meaningful use. *N Engl J Med* (2011):855-858.

Address for correspondence

Dr S.B. Gogia, SATHI, 28/31 Old Rajinder Nagar, New Delhi 110060, INDIA; Email gogia7@gmail.com

Comparative Analysis of Geriatric and Adult Drug Clinical Trials on ClinicalTrials.gov

Zhe He^a, Aisha Langford^b

^a School of Information, Florida State University, Tallahassee, FL, USA

^b Department of Population Health, New York University, New York, NY, USA

Abstract

Clinical trials generate gold standard medical evidence, but are often criticized for the lack of population representativeness. We performed a comparative meta-analysis of drug trials that focus on older adults (≥ 65 years old) and adults (18 – 64 years old). The major finding is that a higher percentage of geriatric drug trials were terminated or withdrawn than that of adult drug trials.

Keywords:

Clinical Trials; Health Disparities; Geriatric Research

Introduction

Clinical trials, especially randomized controlled trials, are widely regarded as gold standard medical evidence. However, they are often criticized for the lack of population representativeness and poor generalizability. Older adults are often underrepresented in clinical trials on various disease domains due to their multiple chronic conditions. In this study, we performed a comparative analysis of drug trials that focus on geriatric and adult populations based on the trial summaries collected from ClinicalTrials.gov.

Methods

We have previously built a relational database of clinical study summaries on ClinicalTrials.gov called COMPACT [1], which includes both trial metadata (e.g., intervention type, intervention name, study phase, study design) and structured eligibility criteria. We downloaded the MeSH-based medical conditions annotation file of the AACT (Aggregate Analysis of ClinicalTrials.gov) database (version: March 27, 2015) [2], developed by US Food and Drug Administration and Duke University. From the COMPACT database [1], we found 654 drug trials that recruited patients ≥ 65 years old only with a start date between January 2005 and September 2016. We analyzed their medical conditions, primary purpose, endpoint classifications and overall status. For the trials that were withdrawn or suspended, we further analyzed the reasons for their withdrawals. We compared the results with adult drug trials that only recruited patients between 18 and 65 years old (minimum age ≥ 18 , maximum age ≤ 65).

Results and Discussion

Top 10 medical conditions of drug trials that focus on older adults included Alzheimer's disease ($n=22$), hypertension ($n=19$), delirium ($n=18$), breast cancer ($n=16$), multiple myeloma ($n=14$), acute myeloid leukemia ($n=13$), non-small

cell lung cancer ($n=12$), osteoporosis ($n=10$), depression ($n=10$), and dementia ($n=10$).

Table 1 shows the comparison of drug trials that recruited older adults and younger adults. A higher percentage of geriatric drug trials were withdrawn or suspended than adult drug trials. The major reasons of termination for the 59 geriatric drug trials were slow accrual (45.8%).

Table 1—Comparison of Geriatric and Adult Drug Trials

Descriptor	Older Adults Drug Trials (N=654)	Adults Drug Trials (N=18,829)
<i>Primary Purpose</i>		
Treatment	499 (76.3%)	12,548 (66.7%)
Prevention	78 (11.9%)	1,066 (5.7%)
Basic Science	19 (2.9%)	2,274 (12.1%)
Diagnostic	12 (1.8%)	321 (1.7%)
Supportive Care	7 (1.1%)	236 (1.3%)
Health Service	3 (0.5%)	131 (0.7%)
Screening	2 (0.3%)	59 (0.3%)
N/A	34 (5.2%)	2,189 (11.6%)
<i>Endpoint Classification</i>		
Safety/Efficacy	327 (50.0%)	5,477 (29.1%)
Efficacy	162 (24.7%)	3,455 (18.4%)
Safety	57 (8.7%)	2,266 (12.0%)
Pharmacodynamics	12 (1.8%)	647 (3.4%)
Pharmacokinetics	10 (1.5%)	2,492 (13.2%)
Bio-availability	4 (0.6%)	458 (2.4%)
Bio-equivalence	2 (0.3%)	900 (4.5%)
<i>Overall Status</i>		
Completed	287 (43.9%)	12,715 (67.5%)
Recruiting	182 (27.8%)	2,850 (15.1%)
Active, not recruiting	64 (9.8%)	869 (4.6%)
Terminated	59 (9.0%)	971 (5.2%)
Not yet recruiting	30 (4.6%)	813 (4.3%)
Withdrawn	22 (3.4%)	411 (2.2%)
Enrolling by invitation	9 (1.4%)	126 (0.7%)
Suspended	1 (0.2%)	74 (0.4%)

References

- [1] He, Z., S. Carini, T. Hao, I. Sim, and C. Weng, A Method for Analyzing Commonalities in Clinical Trial Target Populations. *AMIA Annu Symp Proc*, 2014. 2014: p. 1777-86.
- [2] Tasneem, A., Aberle, L., Ananth, H., et al. The database for aggregate analysis of ClinicalTrials.gov (AACT) and subsequent regrouping by clinical specialty. *PLoS One*. 2012;7(3):e33677.

Address for correspondence

Zhe He, PhD. School of Information, Florida State University, 142 Collegiate Loop, Tallahassee FL, USA 32306.
Email: zhe.he@cci.fsu.edu

Generation of openEHR Test Datasets for Benchmarking

Samar El Helou^a, Tuukka Karvonen^a, Goshiro Yamamoto^b, Naoto Kume^c, Shinji Kobayashi^c, Eiji Kondo^d, Shusuke Hiragi^b, Kazuya Okamoto^b, Hiroshi Tamura^b, Tomohiro Kuroda^b

^a Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto, Japan

^b Division of Medical Information Technology & Administration Planning, Kyoto University Hospital, Kyoto, Japan

^c Department of Electronic Health Records, Graduate School of Medicine, Kyoto University, Kyoto, Japan

^d Department of Gynecology and Obstetrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Abstract

openEHR is a widely used EHR specification. Given its technology-independent nature, different approaches for implementing openEHR data repositories exist. Public openEHR datasets are needed to conduct benchmark analyses over different implementations. To address their current unavailability, we propose a method for generating openEHR test datasets that can be publicly shared and used.

Keywords:

Electronic Health Record; Benchmarking

Introduction

OpenEHR is a technology-independent, open-source specification for Electronic Health Records' (EHR) architecture adopting a two-level modeling approach [1]. The choices of database technologies and approaches for repository implementations are left for the developers. Benchmarking performance of different repository implementations in different use-case scenarios is needed [2]. Usually, benchmark analyses include the comparison of query response times and thus require access to shared openEHR datasets, often unavailable due to strict medical privacy laws.

The effectiveness of a benchmarking dataset is affected by its level of accessibility, realism and evaluation capabilities. In the case of openEHR benchmarking datasets, the structure of the data is constrained by the reference model and archetypes' definitions, thus the evaluation capabilities can be defined and artificially simulated. As for realism, some could be potentially sacrificed in favor of accessibility in cases where real data is difficult to come by.

This work provides a method to generate open application-specific openEHR test datasets. The resultant datasets should comply with openEHR's information and archetype models and allow queries applicable in real world scenarios.

Methods

First, we identified the clinical concepts involved in a pregnancy home-monitoring application and determined a list of realistic data entries. Next, we mapped the clinical concepts to openEHR archetypes available in the openEHR Clinical Knowledge Manager (CKM) and created data value sets corresponding to the possible data entries. We applied an Object Relational Mapping (ORM) approach to design a relational schema allowing the persistence of the required archetypes over classes from the openEHR Reference Model.

We created data generation plans using the archetypes' structure and data value sets, as shown in Figure 1. The plans were executed using Microsoft Visual Studio 2010 to populate an SQL Server database. Finally, we identified application-specific search scenarios for which we formulated and executed SQL queries solely using the archetypes' definitions.

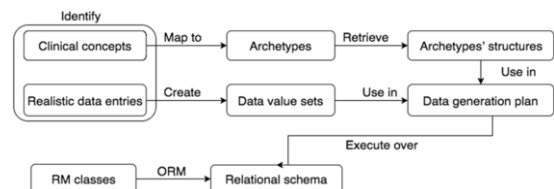


Figure 1 - openEHR dataset generation process

Results

The method was applied to generate datasets simulating a pregnancy home-monitoring repository. A set of seven queries were executed and generated non-empty result sets. Datasets of 10k and 100k records in CSV and JSON formats can be accessed via github.com/samarhelou/data. Cypher queries are also provided to allow dataset import, visualization, and testing in Neo4j, a labeled property graph database.

Conclusions

We proposed and tested a method for generating test openEHR datasets. Future work requires the inclusion of data generation rules to reflect the real distributions of medical cases in the population. The generated datasets will be used to benchmark an openEHR repository implementation using Neo4j.

References

- [1] D. Kalra, T. Beale, and S. Heard, The openEHR foundation, *Studies in health technology and informatics* **115** (2005), 153-173.
- [2] S. Frade, S.M. Freire, E. Sundvall, J.H. Patriarca-Almeida, and R. Cruz-Correia, Survey of openEHR storage implementations, in: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, IEEE* (2013), 303-307.

Address for correspondence

Samar El Helou, email: samar@kuhp.kyoto-u.ac.jp

Mining Associations Between Genes and Clinical Conditions of Breast Cancer by Using Gene Expression Data

Yuefang Hou, Hao Zhang, Han Zhang, Lei Cui

School of Medical Informatics, China Medical University, Shenyang, Liaoning, China

Abstract

Recent studies have identified some genes related to cancer phenotypes, but lack of comprehensive clinical information. Here we present an integrated data mining method to find associations between genes and multiple clinical conditions of breast cancer by using gene expression data from Gene Expression Omnibus. The associated modules of genes and clinical conditions were built and some potential related genes were suggested.

Keywords:

Data Mining; Gene Expression Profiling; Breast Neoplasms

Introduction

High-throughput data are used for establishing expression patterns to recognize specific phenotype, predict the outcome of disease, and promote the development of precision medicine. Recent studies have explored the association between genes and clinical conditions, but lack of comprehensive clinical information. We present an integrated data mining method with gene expression data and clinical annotation of breast cancer. The associated modules of genes and conditions are built for management of cancer.

Methods

Gene expression data of breast cancer in Gene Expression Omnibus GSE2109 were downloaded and organized into gene-sample expression matrix. The data were normalized by logarithmic transformation, mean value centerization for each gene. We selected differentially expressed genes among samples and got 616×358 gene-sample expression matrix. On the other hand, 72 clinical conditions, including personal attribute, behavior, medical history, pathological stage, grade, ER, PR, HER2 and histology, et al, were extracted from the sample descriptions and then standardized to build a 358×72 sample-condition matrix. The above two matrices were multiplied to get a 616×72 gene-condition matrix. Then the association modules between genes and conditions were extracted with biclustering analysis. Both GO and KEGG pathway enrichment analyses were applied to each cluster of genes. We analyzed the biological mechanism of genes to predict potential relations between genes and conditions of breast cancer.

Results

Twenty gene clusters were built with biclustering. The module map shows that 7 gene clusters are correlated to the stage, grade, ER, PR, HER2 and histology of cancer. Another 7 gene clusters are associated with different personal attribute, behavior and medical history. Some associations, such as

LCN2, CDH1, GATA3 and different clinical conditions, are consistent with the prior knowledge in OMIM database. Moreover, we found several potential associations. For example, the expression level of 16 genes in cluster1 (including CDH1) may coordinately affect grade, histology and ER expression of breast cancer. The genes expression of cluster19 are associated with tobacco, alcohol and oophorectomy.

Conclusion

Through the data mining method we developed, the possible associations between genes and clinical conditions of breast cancer are summarized. Our analysis presents multiple research directions for diagnostic, prognostic and therapeutic studies.

Acknowledgements

This work was supported by the Department of Education of Liaoning Province (LR201606).

Address for correspondence

Lei Cui, Email: lcui@cmu.edu.cn

Comparing Different Adverse Effects Among Multiple Drugs Using FAERS Data

Jing Huang^a, Xinyuan Zhang^{b*}, Jingcheng Du^b, Rui Duan^a,
Liu Yang^c, Jason H. Moore^a, Yong Chen^{a#}, Cui Tao^{b#}

^a University of Pennsylvania, Philadelphia, PA, USA

^b University of Texas Health Science Center at Houston, Houston, TX, USA

^c Mayo Clinic, Jacksonville, FL, USA

Abstract

US Food and Drug Administration (FDA) Adverse Event (AE) Reporting System (FAERS) is a major source of data for monitoring drug safety. However, there is not general procedure to systematically compare drugs group. We present a statistical method, which can effectively identify significant differences in AE rates among drugs and estimates the differences in age and gender distributions.

Keywords:

Adverse Drug Reaction Reporting Systems; Data Mining.*

Introduction

Post-marketing surveillance is critical to ensure long-term safety, study rare adverse events (AE), and adverse reactions [1]. The US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) is a post-marketing surveillance program seeking voluntary inputs on AEs to monitor drug safety [2]. We propose an statistical pipeline that provides interpretable results and is convenient to implement. It systematically compares the differences between groups of drugs for specific AEs and can be used to select candidate AEs that are potentially important for future investigation. The method can also be used for other types of reporting systems. We use FAERS data to compare three Hepatitis C therapies.

Methods

The procedure was carried out in three steps.

Step I: Descriptive Analysis and Visualization

Compare the total number of adverse events distribution between the three therapies using side-by-side histogram plots.

Step II: Difference in AE Rates Among Groups

For each type AE type, compare the reporting rates among the different treatment groups using the chi-squared test and select AEs with significantly different reporting rates by therapy group.

Step III: Quantify and Visualize Risk Factor Effect Size

For each selected AE from step II, investigate whether the difference in reporting rates can be explained by differences in demographics by comparing the effect sizes before and after adjusting for these variables.

Results

The FAERS data contains the top 30 most frequently self-reported AEs of the three Hepatitis C therapies: 28,192 patients were in Therapy A, 5,035 in Therapy B, and 7,820 in Therapy C (Therapy A and B drugs). In step I, we observed that Hepatitis C therapy C caused more AEs than B and C. In step II, the chi-squared test significantly different rates of AEs among the three therapy groups, after Bonferroni correction, except for "arthralgia". In step III, we found that for the 30 most frequently reported AEs, the adjusted odds ratio was less than the unadjusted odds ratio for 15 AEs and was greater than the unadjusted odds ratio for 15 AEs. The difference between the adjusted and unadjusted odds ratio attributable to age and gender ranged from 0.2% to 14.8%.

Conclusion

In this paper, we provided a statistical procedure to compare the difference between AEs among multiple drugs using FAERS data. We are currently developing an R package, "AETools", to semi-automate the proposed pipeline for statistical analysis and visualization. Such a procedure, including the R codes, is useful for pharmacoepidemiological studies and will be made publicly available.

Acknowledgements

Research was partially supported by the National Library Of Medicine and National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R01 LM011829, R01 AI116794 and R21 LM012197, and the support from the UTHealth Innovation for Cancer Prevention Research Training Program Pre-doctoral Fellowship (Cancer Prevention and Research Institute of Texas grant # RP160015). Also thank Dr. Jun Xu for his help.

References

- [1] V. Suvama, Phase IV of drug development, *Perspect Clin Res* 1 (2010):57-60.
- [2] FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS): <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/> (2016).

Address for correspondence

Yong Chen: ychen123@upenn.edu

Cui Tao: cui.tao@uth.tmc.edu

* These authors contribute equally

Corresponding authors

Usability Evaluation of NLP-PIER: A Clinical Document Search Engine for Researchers

Gretchen Hultman^a, Reed McEwan^b, Serguei Pakhomov^c,
Elizabeth Lindemann^d, Steven Skube^d, Genevieve B. Melton^{a,d}

^a Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, USA

^b Academic Health Center – Information Systems, University of Minnesota, Minneapolis, Minnesota, USA

^c College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, USA

^d Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA

Abstract

NLP-PIER (Natural Language Processing – Patient Information Extraction for Research) is a self-service platform with a search engine for clinical researchers to perform natural language processing (NLP) queries using clinical notes. We conducted user-centered testing of NLP-PIER's usability to inform future design decisions. Quantitative and qualitative data were analyzed. Our findings will be used to improve the usability of NLP-PIER.

Keywords:

Information Storage and Retrieval; Natural Language Processing; Evaluation Studies.

Introduction

NLP-PIER was created to provide an accessible solution through a search interface to clinical researchers interested in or requiring access to clinical NLP capabilities of clinical documents [1]. The system has two interfaces: a full text search interface and a custom interface for searching Unified Medical Language System (UMLS) concepts [2]. We were interested in understanding potential design opportunities and user acceptance of NLP-PIER and to more broadly understand the needs of clinical researchers when using a self-service NLP tool.

Methods

This study was conducted at University of Minnesota (UMN) as part of its broader clinical and translational science research platform. We designed standard tasks to use NLP-PIER and asked clinical researcher participants (n=11) to complete these tasks, usability instruments (system usability scale (SUS) [3] and NASA-Task Load Index (NASA-TLX) survey [4]), a brief interview, and exit questionnaire. Time on task, task completion percentage, and survey results were assessed. Interviews were transcribed and coded for themes

Results

For the full text search interface, questionnaire scores were 69.4 (19.8) (SUS) and 18.8 (5.7) (NASA TLX) and for the Concept search interface scores were 66.1 (32.4) and 21.8 (7.7). Average time on task and task completion varied widely. In interviews, all participants expressed that NLP-PIER was easy to use and would be useful in their work.

Conclusion

End user testing of NLP-PIER identified a number of usability challenges and several solutions. Our study also demonstrated that substantial variation exists between different users. Overall, our findings illustrate the importance of incorporating user testing and feedback in the design process.

Acknowledgements

This research was supported by the Agency for Healthcare Research & Quality (#R01HS022085 (GM)) and National Institutes of Health (#R01LM011364 (GM), #R01GM102282 (SP), #8UL1TR000114 (Blazar)).

References

- [1] R. McEwan, G.B. Melton, B.C. Knoll, Y. Wang, G. Hultman, J.L. Dale, T. Meyer, S.V. Pakhomov, NLP-PIER: a scalable natural language processing, indexing, and searching architecture for clinical notes, In: *AMIA Summits on Translational Science Proceedings*, 2016 (2016):150.
- [2] US National Library of Medicine. Unified Medical Language System (UMLS), Available from: <https://www.nlm.nih.gov/research/umls/html>.
- [3] J. Brooke, SUS-A quick and dirty usability scale, *Usability Eval Indust* **189** (1996):4-7.
- [4] S.G. Hart, NASA-task load index (NASA-TLX); 20 years later. In: *Proceedings of The Human Factors and Ergonomics Society Annual Meeting*, 2006, pp. 904-908.

Address for correspondence

Genevieve B. Melton, MD, PhD
Institute for Health Informatics and Department of Surgery
420 Delaware Street SE, Mayo Mail Code 450
Minneapolis, MN 55455
Tel: + 612-625-7992
E-mail: gmelton@umn.edu

Explore Care Pathways of Colorectal Cancer Patients with Social Network Analysis

Tianyao Huo^a, Thomas J George Jr^b, Yi Guo^a, Zhe He^c, Mattia Prosperi^d, François Modave^a, Jiang Bian^a

^a Health Outcomes & Policy, University of Florida, Gainesville, Florida, USA,

^b Hematology/Oncology, University of Florida, Gainesville, Florida, USA,

^c School of Information, Florida State University, Tallahassee, Florida, USA,

^d Epidemiology, University of Florida, Gainesville, Florida, USA.

Abstract

Patients with colorectal cancer (CRC) often face treatment delays and the exact reasons have not been well studied. This study is to explore clinical workflow patterns for CRC patients using electronic health records (EHR). In particular, we modeled the clinical workflow (provider-provider interactions) of a CRC patient's workup period as a social network, and identified clusters of workflow patterns based on network characteristics. Understanding of these patterns will help guide healthcare policy-making and practice.

Keywords:

Colorectal Neoplasms, Workflow, Electronic Health Records

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths in the United States. The quicker the patients get diagnosed and treated, the higher the survival rate. Understanding variations of care pathways (i.e., the sequence of clinical encounters) can help improve the workflow efficiency and ultimately quality of care [1]. Many factors may influence these care pathways. This study is thus to explore the characteristics of patients, providers, and the healthcare system that potentially influence the care pathways of CRC patients' workup period using social network analysis (SNA).

Methods

The study cohort is patients diagnosed with CRC identified with ICD-9/10 diagnostic codes. We obtained 1,924 patient records from the University of Florida (UF) Health Integrated Data Repository. The resulting dataset contains demographics, encounters, diagnoses, procedures, and labs of eligible patients between June 2011 and February 2015. The workup period is defined based on procedure codes (i.e., ICD and CPT codes). The starting point of the workup period is the date of the diagnostic procedure (e.g., colonoscopy); and the end point is the date of surgery/chemotherapy. Out of the 1,924 patients, 327 patients' workup periods can be defined (since some patients were not diagnosed or treated at UF Health).

SNA studies the relationships (edges) between a set of actors (nodes). In our study, the nodes are patients' encounter events such as office visits, lab tests, and procedures. The edges depict the relationships between the encounters, such as referrals, orders, and reports, where the edge weight is the time delay between two encounter events. Figure 1 illustrates a care pathway network. We can quantitatively measure network patterns using metrics such as path length.

We then infer clusters of patients that share care pathways (network patterns). We first used random forest (RF) regression to explore the associations between the network metrics and the length of the workup period; and from which, we derived a similarity score for each pair of patients. We used hierarchy agglomerative clustering (HAC) algorithm to

infer clusters of patients based on these similarity measures. We followed best practice in tuning the RF model (e.g., parameter tuning and holdout validation) and conducted the Silhouette analysis to choose the appropriate number of clusters.

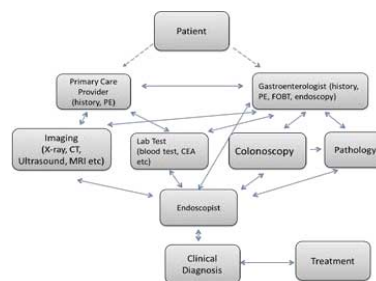


Figure 1. An example of a CRC care pathway network.

Results

The mean workup periods is 98 days with large variations (± 212). On average, patients have 30 (± 43) encounters with 10 (± 9) providers. Using the network metrics as predictors, the trained RF regression model performed well $r^2 = 0.81$. Through Silhouette analysis, we identified 3 clusters of patients using HAC based on the similarity measures.

Table 1 - Descriptive Statistics of the Clusters.

	Cluster 0 (n=110)	Cluster 1 (n=119)
Length (Days)	248.5 (± 315.1)	15.2 (± 10.7)
# of Nodes	16.7 (± 11.6)	7.7 (± 4.7)
Path Length (weighted)	15.7 (± 27.6)	2.1 (± 1.8)
Never Smoker	41.82 %	47.90 %

Table 1 shows some of the characteristics of two clusters. It is clear that these two clusters show distinctive patterns.

Conclusions

Guidelines of CRC, such as those published in the National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology, do exist. However, demonstrated by our study, real world practices are vastly different. Understanding these derivations is critical for reducing time needed for patients to receive appropriate care. SNA is a promising approach in exploring patterns of care pathways.

References

- [1] Li X, Mei J, Liu H, Yu Y, Xie G, Hu J, et al. Analysis of care pathway variation patterns in patient records. *Stud Health Technol Inform.* 2015;210:692-6.

Address for correspondence: Jiang Bian; bianjiang@ufl.edu

Constructing an Open-Access Bio-Signal Repository from Intensive Care Units

Doyeop Kim^a, Sukhoon Lee^b, Tae Young Kim^a, Sanghyung Jin^a,
JaeYeon Park^c, JeongGil Ko^c, Rae Woong Park^{a,d}, Dukyong Yoon^{a,d}

^a Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, South Korea

^b Department of Software Convergence Engineering, Kunsan National University, Kunsan, South Korea

^c Department of Software and Computer Engineering, College of Information Technology, Ajou University, Suwon, South Korea

^d Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, South Korea

Abstract

Bio-signals can be crucial evidence in detecting urgent clinical events. However, until now, access to this data was limited. We aim to construct and provide a new open bio-signal repository with data gathered from more than 40 intensive care unit (ICU) beds. For doing so, we completed the interfacing system with the patient monitors at the target beds and plan to expand this data set to more than 100 ICU beds. Once completed, we plan to publicly open the data to catalyze interesting clinical-event detection research.

Keywords:

Biomimetics, Intensive Care Unit, MIMIC

Introduction

Bio-signals such as electrocardiogram, continuous blood pressure and oxygen saturation can act as evidence in detecting clinical events. The Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II) database provides a diverse set of clinical and bio-signal waveform data from real clinical practice [1]. The database has contributed to many research work from clinical research to biomedical engineering. However, since MIMIC is the only qualified bio-signal database, the researchers using the MIMIC database have suffered from the lack of cross validation data. To address this issue, we plan to construct a new publicly-open bio-signal repository with data collected from ICUs. Our repository is designed to overcome limitations of the MIMIC database, such as the linking issues between clinical and waveform database, data source diversity.

Methods

The subject hospital for data collection is a tertiary teaching hospital in South Korea, with 158 beds spread over eight ICUs equipped with patient monitors from three different vendors, General Electronics, Phillips and Nihon Kohden. Currently, our bio-signal repository interfaces with the Nihon Kohden and Phillips patient monitors. For Nihon Kohden devices, we collect bio-signal data via the Health Level 7 (HL7) protocol using a designated gateway, while Philips devices utilize RS232 connection.

Results

At the current state, we are done interfacing 30 Nihon Kohden devices for beds covering two ICUs. Quantitatively, we collected data from 605 patient from Aug 31 2016 to Apr 6 2017, corresponding to 139,392,966 vital signs samples and 2,910,025 waveform files (10-minute signal data per file).

We plan to finalize the interface process with the Philips devices within the second quarter of 2017. At the end of this phase, we expect a total of 108 connected patient monitors.

Table 1 – data collection status

Collected waveforms	No. Patient	Proportion [*]
ECGII	604	99.83%
Respiration wave	603	99.67%
Oxygen saturation	602	99.50%
Invasive arterial blood pressure	402	66.45%
Central venous pressure	187	30.91%
End-tidal CO ₂	45	7.44%

^{*}Proportion indicates proportion of patients measuring the waveform among entire patients

Discussion

We are constructing the qualified bio-signal repository. It will be linked with clinical database like diagnosis, procedure, or prescription. Furthermore, we aim to open the whole database to the public after anonymization. Different ICU monitoring systems (Nihon Kohden and Philips) generate different format of data. However, since we use a standard data model, the users do not have to consider data sources.

Conclusion

This abstract introduces our efforts for developing and deploying a publicly open bio-signal repository. We envision that our repository can interact with MIMIC as a database for enabling new research in bio-medical informatics.

Acknowledgements

This research was supported by Korea Health Technology R&D project through KHIDI, funded by Ministry of Health & Welfare, Korea (grant number: HI16C0982, HI16C0992).

References

- [1] M. Saeed, M. Villarroel, A.T. Reisner, G. Clifford, L.W. Lehman, G. Moody, T. Heldt, T.H. Kyaw, B. Moody, and R.G. Mark, Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database, *Critical Care Medicine* **39** (2011), 952-960.

Address for correspondence

Dukyong Yoon, MD, PhD

Department of Biomedical Informatics, Ajou University School of Medicine, 206, World cup-ro, Yeongtong-gu, Suwon, South Korea
Tel: +82-31-219-4476 / Fax: +82-219-4472

E-mail: d.yoon.ajou@gmail.com

Depression, Impulse Control Disorder, and Life Style According to Smartphone Addiction

E. Y. Kim^a, S. W. Joo^b, S. J. Han^a, M. J. Kim^a, S. Y. Choi^a

^a Department of Health Care Administration, Seoyoung University, Paju-si, Gyeonggi-do, Republic of Korea,

^b ACE Hospital, Ansan-si, Gyeonggi-do, Republic of Korea

Abstract

We examined depression, impulse control disorder, and life style by degree of smartphone addiction. Chi-square tests and ANOVA were used to identify significant variables. CART was used to generate a decision making diagram of variables affecting smartphone addiction. The severe smartphone addiction group had rates of depression and impulse control disorder than the initial smartphone group.

Keywords:

Smartphone; Behavior, Addictive; Depressive Disorder

Introduction

Smartphones are convenient, many people are using them resulting in tolerance, addiction or difficulties in daily life, depression, learning disability, and impulse control disorders. Smartphone addiction is a serious problem [1,2], especially in the field of information technology [3]. We aimed to predict factors affecting smartphone addiction.

Methods

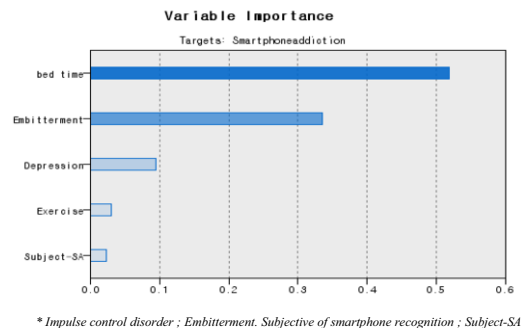
University students (n=132) answered questions about: smartphone addiction, depression, impulse control disorder, and lifestyle. Participants were classified into smartphone addiction groups (normal/suspected/serious), and depression and impulse control disorder groups (normal/poor/serious). Chi-square and ANOVA were used to identify significant variables. CART was used to generate a decision making diagram of variables affecting smartphone addiction.

Results

Men were more addicted to smartphones than women ($p=0.029$). Those in the more severe smartphone addiction groups slept less than 6 hours a night ($p<0.001$) and exercised less ($p=0.011$) than those in the less severe addiction groups (Figures 1 and 2). Severe smartphone addiction groups had more depression ($p=0.005$) and impulse control ($p<0.001$) than the initial addiction group ($p=0.005$). The initial smartphone addiction group had higher impulse control than the normal addiction group ($p=0.002$). Those in the more severe smartphone addiction groups were more likely to objectively indicate having smartphone addiction ($p=0.007$).

Conclusion

Problematic smartphone use increases depression and anxiety. Effectively use requires education to change user preceptions.



* Impulse control disorder ; Embitterment, Subjective of smartphone recognition ; Subject-SA

Figure 1 – Variable importance of CART Modeling



Figure 2 – Expand all Variable of CART Modeling

References

- [1] H.Y. Kang, Development and Validation of the Smartphone Addiction Inventory. *J Kor Psych Assoc* **31** (2012).
- [2] J.T. Oh, J.E. Lee, The 'Smart Life' Revolution and Smart Phone Addiction. *Internet Inf Security* **3** (2012), 21-43.
- [3] D.H. Kim, K.E. Lee, Y.J. Lee, S.H. Cha, S.H. Lim, C.Y. Won, H.M. Kwon, M.A. Han, The Association between Smart Phone Addiction and Attention Deficit Hyperactivity Disorder among Some College Students. *J Kor Soc Matern Child Health*, **17** (2013), 105-112.

Address for correspondence

Eun-Yeob, Kim M.D Ph.D, preferred method of contact email, email; key0227@nate.com, Mobile; +82-10-7115-7900, Office; +82-31-930-9590, Fax; +82-31-930-9599

A Performance Comparison on the Machine Learning Classifiers in Predictive Pathology Staging of Prostate Cancer

Jae Kwon Kim^a, In Hye Yook^b, Mun Joo Choi^b, Jong Sik Lee^a,
Yong Hyun Park^c, Ji Youl Lee^c, In Young Choi^b

^a Department of Computer Science and Information Engineering, Inha University, InhaRo 100, Nam-gu, Incheon, South Korea

^b Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

^c Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Abstract

This study objectives to investigate a range of Partin table and several machine learning methods for pathological stage prediction and assess them with respect to their predictive model performance based on Koreans data. The data was used SPCDB and gathered records from 944 patients treated with tertiary hospital. Partin table has low accuracy (65.68%) when applied on SPCDB dataset for comparison on patients with OCD and NOCD conditions. SVM (75%) represents a promising alternative to Partin table from which pathology staging can benefit.

Keywords: Prostate Cancer, Machine Learning, Pathology staging

Introduction

Machine learning based on prediction of prostate cancer pathological stage is an essential step in a patient treatment [1]. Prediction model determines the treatment that will be applied further. Generally, medical experts use the pathological stage predictions provided in Partin tables to support their decisions. However, Partin tables are based on simple statistical method and built from US data [2]. This study objectives to investigate a range of Partin table and several machine learning methods for pathological stage prediction and assess them with respect to their predictive model performance based on Koreans data.

Methods

The data was collected by the Smart Prostate Cancer Data Base (SPCDB) and gathered records from 944 patients treated with prostatectomy in tertiary hospital [3]. Our study aims at classify the OCD (Organ-Confined Disease; 621 patients) and NOCD (Non Organ-Confined Disease; 323 patients). In addition, a selection of machine learning classifiers including Back Propagation Network (BPN), Support Vector Machine (SVM), Naïve Bayes (NB), Bayesian Networks (BNs), Classification and Regression Tree (CART) and Random Forest (RF) was applied to the same data and using Weka 3.8 software. The classifiers input following variables: PSA, Gleason score (Sum), clinical T stage and positive core count.

Results

We use the 10-cross folds to evaluate performance. Partin tables have low accuracy (65.68%) when applied on SPCDB dataset for comparison on patients with OCD and NOCD conditions. Comparing a range of machine learning classifiers shows that SVM generally outperform other methods. The results of BPN

(73.41%), BN (74.36%), NB (74.79%), CART (70.66%), RF (68.75%) and SVM methods (75%) are overall improved. SVM show higher accuracy than the other methods.

Conclusions

The performance of Partin tables can be described as low to moderate on Korean data. This means that following the predictions generated by Partin tables, Korean patients would received an inappropriate treatment. In addition to demographic differences between Korea and the original US population, the machine learning methodology and in particular Partin table present limitations. Also, SVM represents a promising alternative to Partin table from which pathology staging can benefit.

Acknowledgements

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIP, 2016R1A2B4015922).

References

- [1] O. Regnier-Coudert, J. McCall, R. Lothian, T. Lam, S. McClinton and J. N'Dow, Machine learning for im-proved pathological staging of prostate cancer: a perfor-mance comparison on a range of classifiers. *Artificial intelligence in medicine*, 55(1) (2012), 25-35.
- [2] A.W. Partin, M.W. Kattan, E.N.P. Subong and et al, Combination of prostate-specific antigen, clinical stage and Gleason score to predict pathological stage of localized prostate cancer: a multi institutional update. *JAMA* 277(18) (1997), 1445-1451.
- [3] I.Y. Choi, B.J. Park, B.H. Chung and et. al, Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system. *Prostate Int* 1(2) (2013), 59-64.

Address for correspondence:

In Young Choi, PhD, Department of Medical Informatics, College of Medicine, The Catholic University of Korea, Seoul 137701, Republic of Korea. Email: iychoi@catholic.ac.kr

Avoiding Overfitting in Deep Neural Networks for Clinical Opinions Generation from General Blood Test Results

Youjin Kim, Han-Gyu Kim, Zhun Li, Ho-Jin Choi

School of Computing, KAIST, Daejeon, Korea

Abstract

We have used deep neural networks (DNNs) to generate clinical opinions from general blood test results. DNNs have overfitting problem in general. We believe the complex structure of DNN and insufficient data to be the major reasons of overfitting in our case. In this paper, we apply dropout and batch normalization to avoid overfitting. Experimental results show the improvement in the performance of the DNNs.

Keywords:

Hematologic Tests, Neural Networks (Computer), Clinical Decision-Making

Introduction

Blood tests have been used to detect diseases in the body. One of the previous works analyzed a certain type of blood test results and made a prognosis for one disease using the Deep Neural Network (DNN) [2]. In [1], we created a DNN model that analyzes general blood test results and provided clinical opinions for various diseases. However, the DNN in the previous work suffers from overfitting problem. 1.7 million parameters of the complex DNN and insufficient data cause overfitting. To avoid overfitting, we apply dropout and batch normalization. We find optimal hyperparameter values and get the best performance.

Methods

Overfitting, which occurs with the complex structure of DNN and small size of training data, dramatically decreases the generalization performance of DNNs. The overfitting is highly probable to happen in the clinical domain, as clinical data is generally insufficient.

Dropout temporally cuts connections of hidden layers to avoid overfitting. Cutting connections instantaneously reduces DNN complexity. By applying dropout in various manners, we have found out that it works best when dropout is applied to the first layer of the DNN with dropout rate $p=0.32$.

Batch normalization prevents overfitting by normalizing the activations of the hidden layers which help to regularize DNN. By implementing various experiment on batch normalization, we have discovered that the batch normalization should be applied to all hidden layers in order to obtain the best performance.

Results

We conducted clinical opinions generation experiments on the collected dataset of Koreans containing 14,479 instances.

Table 1 implies that applying dropout and batch normalization improves the opinion generation performance with larger F1-measure values by achieving larger precisions and recalls simultaneously compared to the existing DNN. F1-measure of dropout is slightly larger than that of batch normalization, but the difference is not significant, implying that both of them work similarly.

Table 1– Comparison of Experimental Results

	Precision	Recall	F1-measure
DNN [1]	0.904	0.879	0.888
DNN-Dropout	0.909	0.888	0.898
DNN-BatchNorm	0.907	0.883	0.894

Conclusions

We apply dropout and batch normalization to avoid overfitting in the DNN that generates clinical opinions from general blood test results. Experimental results show that dropout and batch normalization improve the performance of the DNN.

In clinical opinion generation, it is more important to detect the abnormal status of the patients than to find out the normality of healthy persons. We will conduct further researches by focusing on such abnormal cases.

Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program, 10052955, Experiential Knowledge Platform Development Research, funded by the Ministry of Trade, Industry & Energy (MOTIE), Korea.

References

- [1] Y. Kim, J. Hyeon, K.-J. Oh, and H.-J. Choi, Medical Prognosis Generation from General Blood Test Results Using Knowledge-Based and Machine-Learning-Based Approaches, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2016, pp. 125-136.
- [2] M.H. Zheng, K.Q. Shi, X.F. Lin, D.D. Xiao, L.L. Chen, W.Y. Liu, Y.C. Fan, and Y.P. Chen, A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network, *Journal of viral hepatitis* **20** (2013), 248-255.

Address for correspondence

Youjin Kim, School of Computing, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea. Phone: +82 10 8923 5573; Email: 117kyjin@kaist.ac.kr

Li2b2-Façade: Simulation of i2b2 Data Warehouse Server and Client for Interaction with Other Systems

Raphael W. Majeed^{a,b}, Tingyan Xu^a, Mark R. Stöhr^b, Rainer Röhrig^a

^a Department of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany,

^b UGMLC, German Center for Lung Research (DZL), Justus Liebig University, Giessen, Germany

Abstract

Since its release in 2004, the i2b2 data warehouse software has become a valuable tool for clinical researchers. Physicians can use its browser-based query frontend intuitively without additional training or reading through documentation. While the i2b2 software describes its API as “REST”, it is neither stateless nor does it follow the common guidelines for RESTful APIs. Thus, interfacing other software with i2b2’s custom RPC-style XML-API is a very cumbersome process. To overcome these issues, we developed a lightweight software abstraction layer “lightweight i2b2 façade” (li2b2-façade).

Keywords:

Systems Integration; Databases, Factual; Information Storage and Retrieval

Introduction

For the German national registry of emergency care (AKTIN)[1], we installed i2b2 data warehouse servers at several emergency departments all over Germany. Interaction with other software requires certain functionality of the i2b2 server to be accessed from other applications. While the i2b2 software describes its API as “REST”, it is neither stateless nor does it follow the common guidelines for RESTful APIs[2]. Thus, interfacing other software with i2b2’s custom RPC-style XML-API is a very cumbersome process.

To overcome these issues, we identified four requirements in our scenario in order to develop a lightweight software abstraction layer “lightweight i2b2 façade” (li2b2-façade):

(1) Authentication: The i2b2 user management needs to be accessible from other software. (2) The i2b2 query history must be manageable from other applications. (3) Other applications need to run queries in i2b2 and retrieve the results. (4) The i2b2 ontology needs to be accessible from other applications.

Methods

The communications between the i2b2 server and the webclient were analyzed in order to identify the HTTP calls needed to run the i2b2 webclient without a server as well as HTTP calls needed to access the i2b2 server without the webclient.

The library and source code is developed in Java 8, using current security standards and technologies.

Results

Communication analysis of i2b2’s “REST” XML-RPC messages yielded a minimum of 15 calls in four i2b2 components “PM”, “ONT”, “CRC” and “WORK”, which we implemented to satisfy the i2b2 webclient’s functionality. An additional 16 calls were implemented in order to allow complete control over user management and query management.

The developed library “li2b2-façade” was able to satisfy our requirements. All source code is open source and available online at GitHub: <https://github.com/li2b2/li2b2-facade>

Discussion

Usual interaction between the frontend user and the i2b2 data warehouse requires a synchronous connection, which may be difficult for some hospital environments. In this case, the li2b2-façade could be used to allow disconnected clients to run queries asynchronously by establishing a transparent middle layer between client and server. This decoupled client/server architecture allows arbitrary additional processing in-between, like e.g. query manipulation, term mapping or even distributing the query to multiple servers.

Conclusion

We were able to develop a lightweight software abstraction layer “lightweight i2b2 façade”: li2b2-façade, which can be used for (a) simplified access of i2b2 content/functionality from external applications, (b) simulation of an i2b2 server or client, as well as (c) separation and injection of additional functionality between i2b2 server and client. In addition to the AKTIN project [1] which motivated our work, the software is also used by two additional Health IT projects.

References

- [1] J. Ahlbrandt, D. Brammen, R.W. Majeed, et al.: Balancing the need for big data and patient data privacy—an IT infrastructure for a decentralized emergency care research database. *Stud Health Technol Inform.* **205** (2014):750-754.
- [2] L. Richardson, and S. Ruby. *RESTful web services*. O'Reilly Media, Inc., 2008.

Address for correspondence

Raphael W. Majeed, raphael.majeed@chiru.med.uni-giessen.de

Asynchronous Query Distribution Between Multiple i2b2 Research Data Warehouses: Li2b2-SHRINE

Raphael W. Majeed ^{a,b}, Mark R. Stöhr ^a, Volker S. Thiemann ^b, Rainer Röhrig ^b and Andreas Günther ^a

^aUGMLC, German Center for Lung Research (DZL), Justus Liebig University, Giessen, Germany

^bDepartment of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany

Abstract

Clinical Data Ware Houses are established sources for research and quality management. The open source data warehouse software i2b2 enjoys good reputation and wide-spread use in the international medical informatics community. We developed a novel infrastructure to allow queries to be distributed asynchronously between i2b2 data warehouses.

Keywords:

Biomedical Research; Information Storage and Retrieval

Introduction

Clinical Data Ware Houses are established sources for research and quality management. The open source data warehouse software i2b2 enjoys good reputation and wide-spread use in the international medical informatics community. Federated searches across different hospitals are possible with the “Shared Health Research Information Network” (SHRINE) [1] software which unfortunately has some limitations for hospitals with rigid IT security policies and countries with strict data protection – most notably requiring direct network access to participating hospitals as well as not allowing per query approvals by hospitals due to the synchronous query connections.

Aim of this project is the development of a software prototype which overcomes the above shortcomings and allows distribution of queries to federated data warehouses asynchronously without the need for direct network connections to the hospitals.

Methods

The development of an asynchronous network of i2b2 data warehouses can be broken down into three parts: (a) distribution infrastructure, (b) central query frontend and (c) integration software for all participating data warehouses.

For the decentralised query distribution infrastructure, we used software from the German emergency care registry [2]. The infrastructure is content-agnostic and supports any format or language for queries as well as query responses.

Central frontend and data warehouse integration is realized with li2b2-façade [3] with focus on minimal integration effort.

Results

We developed a standalone central server application which employs the original i2b2 webclient to formulate queries.

Queries are stored and can be retrieved by data warehouse nodes at any time. Results are transferred back to the server and displayed in the web frontend.

Terminology and logic of the central queries can be adapted to local characteristics via customizable XML transformations.

Each additional SHRINE-node requires an API-Key in the central server configuration file. The distributed query network was successfully tested with 20 virtual data warehouse servers.

All source code is open source and available online at GitHub: <https://github.com/li2b2/li2b2-shrine>.

Discussion

In contrast to the original i2b2 SHRINE project [1], our solution does not require a full data warehouse installation for the central query interface. The server as well as the data warehouse connector run on any system without installation.

The presented software can be easily adapted to allow other data warehouse software to join a network of i2b2 data warehouse servers: Quick evaluation showed that a commercial data warehouse software with a RESTful query interface could be integrated with 200-300 lines of code (source link above).

Conclusion

Our software “li2b2-shrine” allows data warehouse queries to be distributed to multiple i2b2 data warehouses asynchronously without the need for firewall adjustments. Existing i2b2 data warehouses can be interconnected with minimal effort.

References

- [1] G.M. Weber and et al, "The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories." *Journal of the American Medical Informatics Association* **16.5** (2009): 624-630.
- [2] J. Ahlbrandt, D. Brammen, R.W. Majeed and et al, Balancing the need for big data and patient data privacy--an IT infrastructure for a decentralized emergency care research database. *Stud Health Technol Inform.* **205** (2014): 750-754.
- [3] R.W. Majeed and et al.: Li2b2-façade: Simulation of i2b2 data warehouse server and client for interaction with other systems. *Stud Health Technol Inform* (2017).

Address for correspondence

e-mail: raphael.majeed@chiru.med.uni-giessen.de

Deep Diabetologist: Learning to Prescribe Hypoglycemic Medications with Recurrent Neural Networks

Jing Mei^a, Shiwan Zhao^a, Feng Jin^a, Lingxiao Zhang^b, Haifeng Liu^a, Xiang Li^a,
Guotong Xie^a, Xuejun Li^c, Meilin Xu^d

^aIBM Research China, Beijing, China,

^bDepartment of Computer Science, Peking University, Beijing, China,

^cDepartment of Endocrinology and Diabetes, the First Affiliated Hospital, Xiamen University, Xiamen, China,

^dPfizer Investment Co. Ltd., Beijing, China

Abstract

In healthcare, applying deep learning models to electronic health records (EHRs) has drawn considerable attention. This sequential nature of EHR data make them wellmatched for the power of Recurrent Neural Network (RNN). In this poster, we propose “Deep Diabetologist” – using RNNs for EHR sequential data modeling to provide personalized hypoglycemic medication prediction for diabetic patients. Our results demonstrate improved performance compared with a baseline classifier using logistic regression.

Keywords:

Decision Support Systems, Clinical; Neural Networks (Computer); Electronic Health Records

Introduction

In China, over 100 million adults suffer from diabetes, and most Chinese diabetic patients go to hospitals regularly to receive medications. With large numbers of patients and limited number of diabetologists in China, is there any way to learn to prescribe hypoglycemic medications by deep learning from data (so as to provide clinical decision support for general practitioners in community hospitals)?

Methods

We propose to leverage advanced deep learning algorithms for medication learning from Electronic Health Records (EHR). Technically, we first do EHR data preprocessing (cleansing and imputation). As a result, we extract 481 clinical variables, including 350 3-digit ICD-10 codes, 124 lab tests, and 7 previously used drug classes. Then we apply LSTM for RNN implementation, on a machine configured with GPU using Theano backend. We train each LSTM for 20 epochs and to avoid overfitting, we stack a dropout (of 0.5) layer before the last Dense layer of the output vector. Besides, to address the variable-length inputs to RNN, we set a maximal length of 20 for input (if an instance has length less than 20, then we will mask the rest as 0), and a masking layer with mask value of 0. Hidden layers are set up with 64 cells, and the batch size is 32.

Results

For experiments, we use a cohort of 21,796 patients from a regional EHR repository of a tier II city in China (population of about 3.9 million). The cases to be studied

are outpatient hypoglycemic prescriptions after the first observed type 2 diabetes diagnosis in adults. A case is counted per day per patient, and the total number of cases is 620,633. All models are trained on (randomly selected) 80% of the cases and validated on 10%, remaining 10% for testing.

As shown below, we have 3 predictors: Prev., Logistic Regression (LR) and RNN. The Prev. uses the previous medication for the current prediction. LR is the best candidate for a baseline classifier, while our RNN inputs features at every time step and outputs the 7 drug classes. The evaluation measure is AUC (i.e., the area under the curve), and we report the AUC scores for 7 drug classes, respectively. Results of Prev. are quite high, and the performance of LR is slightly improved, while our RNN outperforms Prev. and LR.

AUC	Biguanides	Sulfonylureas	Glinide	TZDs	AGIs	DPP-4	Insulin
Prev.	0.8630	0.8798	0.8426	0.8245	0.8541	0.8823	0.9164
LR	0.8928	0.9210	0.8676	0.8481	0.8305	0.8761	0.8905
RNN	0.9252	0.9428	0.9326	0.9169	0.9104	0.9058	0.9422

Besides, we report the results of accuracy = #Hit/#Sample, where #Hit is the number of cases in which the predicted label equals to the targeted label, and #Sample is the number of samples. Moreover, we separately report results for those head cases, tail cases, and all cases – the head case means the first medication of a patient’s sequence, and the tail case means the last medication of a patient’s sequence, while the average is calculated for all cases of a patient’s sequence. The head accuracy of Prev. is not available, since there is no previously prescribed medication to count in for the head case.

Accuracy	Head	Tail	Average
Prev.	--	0.6772	0.6456
LR	0.2963	0.6986	0.6693
RNN	0.3417	0.7060	0.6733

Address for correspondence

Jing Mei, PhD. IBM Research China, Building 19, Zhongguancun Software Park, Haidian District, Beijing, China, 100193.
Email: meijing@cn.ibm.com

System Development Aiming Efficient Recording of Pedigree Information and Database Construction for Genetic Counseling Clinic

Takuya Nagasawa^a, Mitsuhiro Azuma^a, Michihiko Aki^a, Masaki Yamada^a, Mineko Ushijima^d, Noriko Tanabe^b, Takeshi Nakajima^c, Kokichi Sugano^b, Hiromi Sakamoto^d, Teruhiko Yoshida^d

^a FUJITSU LIMITED, Tokyo, Japan

^b Department of Genetic Medicine and Services, National Cancer Center Hospital, Tokyo, Japan,

^c Department of Endoscopy, National Cancer Center Hospital, Tokyo, Japan,

^d Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan

Abstract

In the collaborative research to support genomic medicine, we aim to improve the efficiency of operation such as prevention of hereditary cancer syndromes. In the present work, we built a prototype system to record a pedigree chart, clinical and genetic information on individuals during a genetic counseling session. In a mock examination, we were able to draw the pedigree chart of four generations in about four minutes and to record necessary information without disrupting conversation with counselees.

Keywords:

Hereditary Cancer Syndrome; Genetic Counseling; Pedigree

Introduction

FUJITSU Ltd. and the National Cancer Center (NCC) have carried out collaborative research for the development of ICT to provide the support required for introducing genomic medicine in a cancer clinic. As one of the approaches to this, we commenced prototyping and verification of the system that is required for supporting genetic counseling for consultation, diagnosis, and prevention of hereditary cancer syndromes.

Methods

We are compiling information such as transmission of genetic traits and medical history to a pedigree chart for the prognosis of the development of heredity tumors during a genetic counseling session. So, the major requirements are as follows: 1) A pedigree chart can be created easily and quickly during a conversation with the counselee, 2) attribute information and the contents of consultation can be recorded easily at the same time as that in the pedigree chart, 3) information can be managed, retrieved, and accessed in pedigree units.

Of these requirements, we built a prototype system—Pedigree Chart Editor and Pedigree Viewer. The entry format of the pedigree chart complies with that of the National Society of Genetic Counselors (NCGS) [1].

The system was tested by four members such as doctors in charge of outpatients of genetic counseling to mainly evaluate the operability of Pedigree Chart Editor. By conducting mock genetic counseling with a model pedigree chart, the time required for creating a pedigree chart was measured.

Results

All four members evaluated the system with regard to its operability and suitability for the support of genetic counseling and efficiency improvement. The members were able to draw a pedigree chart of four generations only within the kinship in about four minutes. On the other hand, recording of attribute information required about fifteen minutes in total. In the mock genetic counseling, the members could input information without interrupting the conversation.

Discussion

In the development of the system, particular attention was given to “concise operability without interrupting a conversation”. A new data input interface, “Float Menu”, delivered the required result. This function realized the idea of displaying mental options on the input section with minimum pointer movements from the current input section. This was particularly useful for having concise operability and improving the input performance.

Recording of attribute information that involves input of characters and numeric digits requires a longer time than inputting kinships. To enhance the objectives of the system, it is necessary to introduce a menu for the frequently input patterns and to review the operation structure.

Conclusions

In this research, we built a prototype system that can support genetic counseling. The result of testing showed that the system allows for the 1) recording of information without disruption of conversation with counselees and 2) efficient digitization of pedigree charts.

References

- [1] Robin L. Bennett et al. Standardized Human Pedigree Nomenclature: Update and Assessment of the Recommendations of the National Society of Genetic Counselors. *Journal of Genetic Counseling*, 2008, 17, 424–433.

Address for Correspondence

Takuya Nagasawa: nagasawa_takuya@jp.fujitsu.com

Estimating Sample Size for a Feasibility Study of Computer-Assisted Input Support to EDC

Naoki Nishimoto^a, Naomi Tamura^b, Jun Kunikata^a, Sumiko Akahori^c,
Tomoaki Sogo^c, Yuma Tani^d, Hideto Yokoi^{a,c}

^a Clinical Research Support Center, Kagawa University Hospital, Kita-gun, Japan

^b Graduate School of Health Sciences, Hokkaido University, Sapporo, Japan

^c Department of Medical Informatics, Kagawa University Hospital, Kita-gun, Japan

^d Graduate School of Medicine, Kagawa University, Kita-gun, Japan

Abstract

The aim of this study was to estimate the sample size for the assumed feasibility study of the computer-assisted input support on a clinical trial. More than 1,500 observations were required for the feasibility study with assumed settings. Further study was required for more efficient research design.

Keywords:

Sample size, electric data capturing, hospital information systems

Introduction

In clinical trials, clinical research coordinator (CRC) inputs clinical information to an electronic data capturing (EDC) on their daily work. Direct data transfer was required for precision and reducing the labor intensive work, however, electronic health record (EHR) has been disconnected from the internet for the security reason. EHR data was standardized by HL7 format in almost all the university hospital of Japan. Computer assisted input support on clinical trials was not studied well so far, we are developing the tool for the work.

We are planning the feasibility study to evaluate the performance of the tool. In the software validation, sample size estimation was less implemented. Binary outcome, such as success or fail, was well interpretable however, it was difficult to scoring the software performance. Zhang et al. proposed the sample size estimation method for the paired binary outcome for this problem.

In our research, the aim of this study was to estimate sample size for the assumed feasibility study of the computer-assisted input support for the clinical trial.

Methods

We assumed the following scenario:

1. CRC or other clinical staffs, such as nurses, pharmacists, laboratory technologist
2. Computer assisted input vs manual input
3. Cross-over design with paired binary outcomes
4. Some missing was assumed from EHR

At first, we calculated the sample size based on the exact McNemar test using SAS 9.4 (SAS Institute Inc., Cary NC) with significance level 5% and power 80%. Then, we used Zhang's method for the sample size calculation and simulation with missing data[1].

Results

When the difference of paired proportion was 0.05 (P1-P0) and the correlation was increased from 0.0 to 0.5 by 0.1, then we obtained the sample size from 1537 to 770. Then, we set the correlation was 0 in Zhang's method with missing proportion increased from 0.01 to 0.1 by 0.01 in pre- and post treatment. The estimated sample size was shown in the table.

Table 1 Sample size estimation results.

Correlation	exact McNemar test	Correlation	Missing proportion	Zhang's method
0.0	1,537	0.0	0.01	1,554
0.1	1,400	0.0	0.02	1,570
0.2	1,250	0.0	0.03	1,586
0.3	1,100	0.0	0.04	1,603
0.4	920	0.0	0.05	1,619
0.5	770	0.0	0.06	1,637
		0.0	0.07	1,654
		0.0	0.08	1,672
		0.0	0.09	1,690
		0.0	0.10	1,709

Conclusions

A feasibility study of computer-assisted input in a clinical trial required over 1500 inputs. It depends on the correlation of paired input by a CRC. In software validation, sample size estimation was less implemented so far. Binary outcome was well interpretable and we should note larger input trial was required for the performance evaluation.

References

- [1] S. Zhang, J. Cao, and C. Ahn, Inference and sample size calculation for clinical trials with incomplete observations of paired binary outcomes, *Stat Med* (2016).

Address for correspondence

Naoki Nishimoto, Ph.D.

Clinical research support center, Kagawa University Hospital

1750-1 Ikenobe, Miki-cho, Kita-gun, Kagawa Prefecture, 761-0793 Japan

Phone: +8187-891-2135, Email: nnishimoto@med.kagawa-u.ac.jp

Extracting Predictive Indicator for Prognosis of Cerebral Infarction Using Machine Learning Techniques

Yasunobu Nohara^a, Koutarou Matsumoto^b, Naoki Nakashima^a

^a Medical Information Center, Kyushu University Hospital, Fukuoka, Japan

^b Saiseikai Kumamoto Hospital, Kumamoto, Japan

Abstract

Identifying important predicative indicators for prognosis is useful since these factors help for understanding diseases and determining treatments for patients. We extracted important factors for prognosis of cerebral infarction from EHR. We analyzed EHR data of 1,697 patients with 1,602 variables using gradient boosting decision tree. Extracted factors include not only well-known factors such as NIHSS but also new factors such as albumin-globulin ratio.

Keywords:

Prognosis; Cerebral Infarction; Machine Learning

Introduction

It is important to know predictive indicators for prognosis since these indicators are helpful to understand diseases and to decide treatments for improving patients' outcome. Clinical information is entered into EHR systems by providers and ancillary systems such as radiology and laboratory information systems. In this paper, we extract predicative indicators for prognosis of cerebral infarction from EHR using machine-learning techniques.

Methods

A total of 1,849 patients were admitted to Saiseikai Kumamoto Hospital for cerebral infarction from October 2011 to October 2016. These patients were entered into a clinical pathway for mild cerebral infarction. We excluded patients who had incomplete data. Finally, we selected 1,697 patients for the study.

Objective variable

The modified Rankin Scale (mRS) is a scale for measuring the degree of disability in the daily activities of people who have suffered a stroke. The discharge mRS ≥ 3 , that means patients need assistance for daily activities, is typically defined as poor prognosis and we use this definition as the objective variable.

Explanatory variables

We used all EHR data that were available by the first day of admission as explanatory variables. We extracted 1,602-variables from five information sources: hospital discharge summary, discharge summary of neurology department, nurse observations and clinical inspection, and nutrition assessment.

Analysis methods

Since non-linear relationship exists in medical data and feature importance helps to interpret results, we use gradient boosting decision tree (GBDT) [1], one of the tree-based machine learning algorithms for making prediction model. GBDT also in-

cludes feature selection mechanism and is robust with many explanatory variables. A variable importance shows the strength of association between explanatory variables and the outcome; but does not show how they relate. We use a partial dependence plot (PDP) [1], which is useful to show the effect of the target explanatory variable for outcome [2].

Results and Conclusions

We generate a prediction model for poor prognosis using GBDT and extract important factors. The area under curves (AUCs) of the predictor is 0.839 and means excellent performance for prediction. Extracted important factors, of course, include many widely known predictive factors for prognosis of cerebral infarction such as pre-symptomatic mRS, NIHSS and age. Although not widely known, albumin-globulin ratio (A/G ratio) is extracted as the seventh important predictive factor. Figure 1 shows PDP for poor prognosis risk against A/G ratio. Patients with high A/G ratio (>1.5) belong to low risk group. This result is consistent with latest study by Tomasz et al. [3]. Machine learning techniques are useful to extract important factor for prognosis and understand how they relate to outcome.

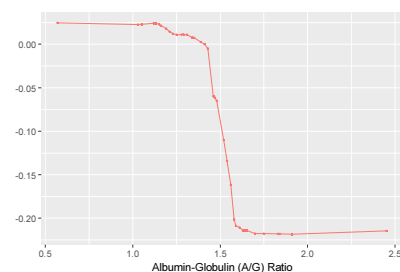


Figure 1— Partial dependence plot that shows relationship between albumin-globulin ratio and risk of poor prognosis

References

- [1] J.H. Friedman, Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics* **29**(5) (2001), 1189-1232.
- [2] Y. Nohara, Y. Wakata and N. Nakashima, Interpreting Medical Information Using Machine Learning and Individual Conditional Expectation, *Proceedings of the 15th World Congress on Medical and Health Informatics (MedInfo2015)*, p.1073, 2015.
- [3] T. Dziedzic, A. Slowik and A. Szczudlik, Serum Albumin Level as a Predictor of Ischemic Stroke Outcome, *Stroke* **35** (2004), e156-e158.

Address for correspondence

Email address: y-nohara@info.med.kyushu-u.ac.jp

Identifying Patients' Smoking Status from Electronic Dental Records Data

Jay Patel^{a,b}, Zasim Siddiqui^a, Anand Krishnan^a, Thankam Thyvalikakath^{a,c}

^a Indiana University School of Dentistry, Indianapolis, IN, USA

^b Department of Bio-Health Informatics, Indiana University School of Informatics and Computing, Indianapolis, IN, USA

^c Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, USA

Abstract

Smoking is a significant risk factor for initiation and progression of oral diseases. A patient's current smoking status and tobacco dependency can aid clinical decision making and treatment planning. The free-text nature of this data limits accessibility causing obstacles during the time of care and research utility. No studies exist on extracting patient's smoking status automatically from the Electronic Dental Record. This study reports the development and evaluation of an NLP system for this purpose.

Keywords:

Smoking; Dental Records; Clinical Decision-Making.

Introduction

Smoking is a significant risk factor for initiation and progression of oral diseases such as periodontal disease, dental caries, and oral cancers. Therefore, it is crucial for dental clinicians to be aware of patient's current smoking status and tobacco dependency [1]. This can help clinicians to make decisions and to take preventive measures as well as for treatment planning. In addition, this information can be used for conducting large scale research studies such as studies related to the association and correlation of smoking and oral diseases. However, due to free-text nature of the data, the access to this information can cause obstacles during the time of care and can be limiting for research purposes [2]. Historically, manual review is required in order to use this information. However, manual chart review can be labor intensive, expensive, and time-consuming. Natural language processing (NLP) can automatically extract patients' smoking status from these histories to reduce human effort. Currently, no study exists on extracting patients' smoking status automatically from Electronic Dental Record (EDR) data. This study reports on developing and evaluating an NLP system to identify patient's smoking status from free text data in EDR.

Methods

This study was approved by the Indiana University Institutional Review Board (Study #:1611054551) and conducted at the Indiana University School of Dentistry (IUSD). We extracted de-identified clinical notes of patients who underwent oral examination at IUSD between December 31, 2011-January 1, 2012 from the EDR. Next, two clinicians trained in informatics manually reviewed and annotated 555 sentences describing patients' smoking status as a smoker, non-smoker, or past smoker. We performed Cohen's kappa statistical test to find the inter-annotation agreement (IAA)

between the two annotators and any disagreement was resolved through discussion and consensus. We considered this dataset our gold standard and we divided this dataset into training (389) and testing (166) sets. We used the training dataset to develop our NLP algorithm. We evaluated the performance of NLP algorithm by testing its accuracy, precision, and recall on the test set.

Results

We observed an IAA of 92%. Our system achieved high precision (98%), recall (98%), and f measure (99%) in differentiating smokers and non-smokers. We observed 80% precision, recall, and accuracy when classifying patients into past smokers. The reason behind the moderate performance could be due to variations in documentations concerning past smoking status of patients in the EDR.

Conclusion

Our NLP system performed excellent in classifying patients' smoking status into smoker and non-smoker and had moderate performance in classifying past smokers. In the future, we will annotate more smoking histories and increase the size of our training set. In addition, we will capture more variations in documenting past smokers. We will run this algorithm on bigger dataset and use this information for research purposes.

References

- [1] G., Smoking, smokeless tobacco, and alcohol consumption as contributing factors to periodontal disease, *Northwest Dent J* **95** (2016): 37-42.
- [2] DT. Heinze, ML. Morsch, BC. Potter, RE. Sheffer Jr, Medical i2b2 NLP smoking challenge: the A-Life system architecture and methodology, *J Am Med Inform Med* **15**(2008.): 40-43.

Address for correspondence

Thankam Paul Thyvalikakath, DMD, MDS, PhD
Associate Professor & Director of Dental Informatics Core;
Research Scientist, Regenstrief Institute, Inc.
Indiana University School of Dentistry, IUPUI
1121 W Michigan Street, Room DS314
Indianapolis, IN 46202
Email: tpt@iu.edu
Phone: (317) 274-5460.

Discovery of Psychoactive Substance Addiction Patterns Based on Information Mining Engineering

García-Martínez Ramón^{a,b}, Martins Sebastián^a, Bianco Santiago^a, Navas Hernán^c

^a Information Systems Research Group, National University of Lanus, ^b Buenos Aires Scientific Research's Commission, ^c Health Informatics Department. Sanatorium Finochietto, (Buenos Aires, Argentina).

Abstract

The evolution of Medical Data Mining has been possible from use of Health Information Systems. In Argentina, consumption of psychoactive substances is quantified through the National Survey on Prevalence of Psychoactive Substance of the National Institute of Statistics and Census. This paper presents the use of Information Mining Engineering as an alternative approach to identifying novel patterns.

Keywords:

Data Mining, Public Health Informatics

Introduction

Information Mining engineering (IME) covers the processes and methodologies used to order, control and manage the task of finding knowledge patterns in masses of information. Gervilla et al reports a study whose objective is to analyze and quantify the predictive value of different personal, family and environmental variables in cannabis use in adolescence. [1] In this context, our objective is to explore the use of IME as an alternative approach to identify patterns on the consume of psychoactive substance, to obtain qualitative answers to: i) Which are factors that describe whether a person tend to consume or not, marijuana? and ii) Which variables characterize different groups among the addicted people (and their level of incidence), allowing identify vulnerable sectors?

Methods

This research uses: a) *Discovery of Behavior Rules (DBR)*: to identify the conditions that describe a specific outcome in the problem domain, b) *Weighting of Group-Membership Rules (WGMR)*: to identify which are the conditions of membership to each of the classes of an unknown partition "a priori". Involving the use of the combination of algorithms: Self-Organizing Map, Bayesian Networks and Decision Trees. [2]

Results

The database contains 34343 records. For the research question [i], the process DBR is implemented, using "Person consumes Marijuana" as Class Attribute. For question [ii], the process WGMR is implemented, filtering by people who consumes any psychoactive substance. The results are shown in Figure 1 and 2 respectively, describing the most relevant factors between parenthesis.

IF Person knows other people that consume psychoactive Substances	IF Person knows other people that consume psychoactive Substances
AND Person has Drink habits	AND Person has Drink habits
AND Person has smoking habits	AND Person has not smoking habits
THEN Person is likely to consume Marijuana	AND Educational Level is incomplete college – university degree or higher
	THEN Person is likely to consume Marijuana
IF Person doesn't know other people that consume psychoactive Substances	IF Person knows other people that consume psychoactive Substances
THEN Person is not likely to consume Marijuana	AND Person has Drink habits
	AND Person has not smoking habits
IF Person knows other people that consume psychoactive Substances	AND Educational Level is complete high school or lower
AND Person has not Drink habits	THEN Person is not likely to consume Marijuana
THEN Person is not likely to consume Marijuana	

Figure 1. Obtained Rules for Research Question [i]

IF sex is Female	IF sex is Male
AND is older than 38 years old	AND is older than 38 years old
AND has incomplete college – university degree or higher	THEN Group 2 (tranquilizers habits – 0.90)
THEN Group 1 (not significant attributes)	IF sex is Female
	AND is older than 38 years old
IF sex is Female	AND has complete high school degree or lower
AND is younger or equal to 38	THEN Group 2 (anxiolytics habits – 0.78)
THEN Group 3 (anxiolytics habits – 0.86)	IF sex is Male
	AND is younger or equal to 38
	THEN Group 4 (marijuana habits – 0.94)

Figure 2. Obtained Rules for Research Question [ii]

Conclusion

The discovery of psychoactive substance addiction patterns may be the cornerstone to identified state policies. When a information mining engineering protocol is establish, the results of application of state policies may be revealed comparing patterns discovered in different editions of the same survey.

Acknowledgements

In memoriam to Ramón García Martínez. Partially funded by: UNLa-80020160500002LA, National University of Lanus.

References

- [1] E. Gervilla, B. Cajal, and A. Palmer, Quantification of the influence of friends and antisocial behaviour in adolescent consumption of cannabis using the ZINB model and data mining, *Addictive behaviors* **36** (2011), 368-374.
- [2] R. García-Martínez, P. Britos, and D. Rodríguez, Information mining processes based on intelligent systems, in: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, Springer, 2013, pp. 402-410.

Address for correspondence

Hernán Navas (hermanjosnavas@gmail.com)

Bringing Knowledge to Users in One Click: Infobuttons in the Problem List of an EHR

Jose F. Rodriguez, Emilio S. Mole, Luciana Rubin, Sonia E. Benitez,
Carlos M. Otero, Daniel R. Luna, Fernan Quiros

Department of Health Informatics, Hospital Italiano, Buenos Aires, Argentina

Abstract

The infobuttons allows the solving of information needs. In our study, the use of Infobuttons is described, analyzing the number of queries to UpToDate® from the problem list of an Electronic Health Record. There were 26419 requests in 8 months. The highest average use occurred in June. The links to knowledge bases can help to solve information needs, even before they occur.

Keywords: Electronic Health Records, Decision Support Techniques, Information Storage and Retrieval

Introduction

In medical practice it's common for health practitioners to have doubts, especially regarding clinical cases which they deal with on a daily basis, particularly those related to diagnosis and treatment [1]. The availability of knowledge bases, such as UpToDate®, depends on a subscription by the user, and it also requires the accomplishment of a manual search, which presents a certain complexity in order to achieve the expected results [2].

The infobuttons are tools that allow different patient-specific data to be added to a search, allowing the user to obtain results with greater completeness and precision, using HL7 interoperability standards [3]. The purpose of this study was to describe the use of Infobuttons in our Electronic Health Record (EHR).

Methods

A descriptive, observational study was carried out analyzing the Infobuttons usage records in the problem list of the EHR. The data collection was performed from March 21, 2016 until October 31, 2016. The use was determined by each log accessing the tool taking into account the click on Infobuttons to access the response of UpToDate®.

Results

In the analyzed period, a total of 26410 queries were registered to the Infobuttons tool, of which 905 occurred during last 10 days of March, 3855 in April, 3462 in May, 4704 in June, 3685 in July, 3629 in August, 3216 in September, and finally 2963 in October. When we analyzed the tool use, discriminating by day of the week, it was possible to appreciate that the greatest number of requests to Infobuttons were made during Monday, Tuesday, Wednesday and Thursday, accounting for 4614, 5141, 4832, and 4873 interactions, respectively.

The analysis of use, according to schedule, showed that most of requests were performed in the morning hours between 9

am and 12 pm. A peak was detected at 11 am accounting for a total of 2348 requests to Infobuttons with a decrease to 214 at 5 am. Regarding to the daily average, 117.9 consultations per day were registered, while during the first month of the implementation of the tool was maintained in 90.5 queries per day. An increase was detected in the month of June to 156.8 per day, remaining at around 100 daily visits in the remaining months. The problems that more frequently needed the contextual help provided by Infobuttons as a clinical decision support were 'Patient reviewed', 'Preoperative state', 'Clinical Finding', 'Patient currently pregnant', 'High blood pressure', 'Procedure related finding', 'Abdominal pain', 'Postoperative state', 'Nutritional support', and 'Tobacco dependence syndrome'.

Conclusion

In this study, we obtained an understanding of the online resources and the Infobuttons use in a problems list. The availability of these links to knowledge bases can help solve information needs, potentially improving patient care and decision making.

References

- [1] S. Hunt, J.J. Cimino, D.E. Koziol, A comparison of clinicians' access to online knowledge resources using two types of information retrieval applications in an academic hospital setting. *J Med Libr Assoc* **101** (2013), 26–31.
- [2] G. Del Fiol, T.E. Workman, P.N. Gorman, Clinical questions raised by clinicians at the point of care: a systematic review. *JAMA Intern Med* **174** (2014), 710–718.
- [3] J.J. Cimino X. Jing, G. Del Fiol, Meeting the electronic health record "meaningful use" criterion for the HL7 infobutton standard using OpenInfobutton and the Librarian Infobutton Tailoring Environment (LITE). *AMIA Annu Symp Proc* **2012** (2012), 112–20.

Address for correspondence

José Federico Rodríguez, Department of Health Informatics, Hospital Italiano, Juan D. Perón 4190, C1181ACH Buenos Aires, Argentina.
E-mail: josef.rodriguez@hospitalitaliano.org.ar.

Effects of e-Learning and m-Learning on Nursing Care in a Continuing Education Context: An Overview of Mixed Method Systematic Reviews (Protocol)

Geneviève Rouleau^{a,b}, Marie-Pierre Gagnon^{a,c}, José Côté^{b,d}, Emilie Hudson^b, Julie Payne-Gagnon^c, Julien Bouix-Picasso^{d,e}, Carl-Ardy Dubois^d

^a Faculty of Nursing Sciences, Université Laval, Quebec, Canada

^b Research Center of the Centre Hospitalier de l'Université de Montréal, Montréal, Canada

^c Research Center of the CHU de Québec, Québec, Canada

^d Faculty of Nursing Sciences, Université de Montréal, Montréal, Canada

^e Laboratory of Health Education and Practices EA3412, Université Paris 13, Paris, France

Abstract

Continuing education is an imperative for professional nursing. e-Learning is one modality to support education and it has been extensively examined in a nursing academic context. An overview of quantitative, qualitative, and mixed-method systematic reviews were conducted to draw a broad picture of the effects of e-Learning and m-Learning used by registered nurses in a continuing education context.

Keywords:

Students, Nursing, Learning

Introduction

Continuing education (CE) is an imperative for nursing creating opportunities for nurses to acquire knowledge and develop competencies [1]. CE should improve the quality of care and patient health by changing the practice of providers [2]. E-learning represents a potential alternative way to learn and has positive impacts on nurses' knowledge, skills, and satisfaction [3]. E-Learning technologies have been studied extensively for health professional students in an academic context (systematic reviews (n=22)) [4]. We found no overview of reviews focused on e-learning in a CE context for registered nurses (RN). We aim to systematically summarize evidence from qualitative, quantitative, and mixed method systematic reviews regarding the effects of e-learning and m-learning on nursing care. The methods and preliminary results will be presented.

Methods

The Cochrane methodology provides guidance overview preparation [5]. General health sciences (PubMed), nursing (CINAHL), and education electronic databases (ERIC) in addition to systematic reviews (Cochrane Database of Systematic Reviews, Epistemonikos) will be searched. The eligibility criteria are formulated using PICOS (Participants, Interventions, Comparisons, Outcomes, Studies). **P:** RN e-learning/m-learning. **I:** A variety of e-learning/m-learning interventions in a nursing CE context. **C:** comparing e-learning/m-learning with other learning. **O:** Conceptualize how e-learning/m-learning interventions could influence nursing. An organizational model will be used: the Nursing Care Performance Framework [6]. Outcomes include: Nursing resources (e.g., nursing staff supply, working conditions); nursing processes (e.g., work environment, nursing activities, satisfaction); and nursing outcomes (e.g., patient experience, quality of life, empowerment). **S:** Quantitative, qualitative, and mixed method systematic reviews published in English, French

or Spanish from January 1, 2006 will be considered. Four reviewers will: independently screen the reviews to assess their eligibility, summarize the characteristics of reviews, perform data extraction and assess methodological quality. Data will be summarized using a narrative synthesis [7].

Discussion

This overview could elucidate the dimensions of nursing care with the potential to be supported, enhanced or constrained by the use of e- and m-learning in CE efforts.

Acknowledgements

This work is funded by the Quebec Nursing Intervention Research Network.

References

- [1] K. Ross, J et al., Mandatory continuing professional development requirements: what does this mean for Australian nurses, *BMC Nur* **12** (2013), 1-7.
- [2] D. Moore et al., Achieving desired results and improved outcomes: Integrating planning and assessment throughout learning activities, *J Contin Educ Health Prof* **29** (2009), 1-15.
- [3] M. Lahti et al., Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: A systematic review and meta-analysis, *Int J Nurs Stud* **51** (2014), 136-149.
- [4] W. De Caro et al., E-learning and university nursing education: an overview of reviews, *Professioni infermieristiche* **67** (2014), 107-116.
- [5] J. Higgins and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration, 2011.
- [6] C.A. Dubois et al., Conceptualizing performance of nursing care as a prerequisite for better measurement: a systematic and interpretive review, *BMC Nurs* **12** (2013), 7.
- [7] J. Popay et al., Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Final Report, in: *Swindon: ESRC Methods Programme* 2006.

Address for correspondence

Research Center of the Centre Hospitalier de l'Université de Montréal, 850 rue St-Denis, Tour St-Antoine, third floor, Door S03-422, Montreal (QC), Canada, H2X0A9.
genevieve.rouleau.chum@ssss.gouv.qc.ca

Using Human Phenotype Ontology for Phenotypic Analysis of Clinical Notes

Feichen Shen, Liwei Wang, Hongfang Liu

Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

Abstract

Phenotypes are defined as observable characteristics of organisms. To facilitate the translation between genotype and phenotype, Human Phenotype Ontology (HPO) was developed as a semantically computable standardized vocabulary to capture phenotypic abnormalities found in human. In this study, we investigated the use of HPO to annotate phenotypic information in clinical domain by leveraging a corpus of 12.8 million clinical notes created from 2010 to 2015 for 729 thousand patients at Mayo Clinic Rochester campus.

Keywords:

Semantics; Human Phenotype Ontology, Phenotypic Analysis

Introduction

Phenotypes, defined as observable characteristics organisms, have attracted increasing attentions in the area of translational medicine by serving as the connectors between medical experimental findings and clinical practices. As a tool for annotating human phenotypic abnormalities, Human Phenotype Ontology (HPO) [1] has been developed as a controlled vocabulary for phenotypes by integrating phenotype knowledge from medical literatures and disease/gene databases. Here, we used HPO to annotate a large collection of clinical narratives consisting of all clinical notes generated from 2010 to 2015 at Mayo Clinic Rochester campus and assessed the distributional information of HPO terms.

Methods

Our annotation workflow is designed to leverage the Unified Medical Language System (UMLS) and corresponding NLP tool MetaMap. Figure 1 shows the overview of our annotation workflow. It includes two modules: i) the preprocessing module which leverages NLP techniques to extract key words from text; and ii) the semantic annotation module which maps the lexicon entries acquired from text to HPO terms.

Results

Demographic Phenotype Analysis- we grouped age 0-17, 18-35, 36-54 and >55 as children and teenagers, young adults, mid adults, and old adults age groups. 8 of the most popular HPO abnormality categories for each age group are shown in Figure 2. In addition, we found that *Hypergonadotropic hypogonadism*, *Hirsutism*, *Basal cell carcinoma* and *Scarring alopecia of scalp* are most significant phenotypes for male, female, White and Black or Africa patients, respectively.

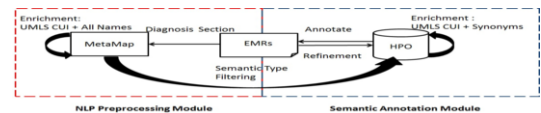


Figure 1—Annotation Work Flow

A Case Study of Phenotypic Analysis- For Wilson's Disease (WD), two case-control studies were designed as follows: 1) Odds ratios of phenotypes were computed for 39 confirmed WD patients compared to 615,488 Non-WD patients; 2) Odds ratios of phenotypes were computed for 39 confirmed WD patients compared to 63 patients suspected for WD. Based on odds ratio, it showed that *Hepatic failure* is the most significant phenotype for study 1 and *Hyperglycemia* is the most popular phenotype for study 2.

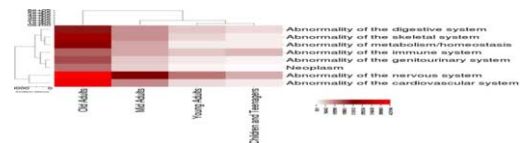


Figure 2—Distribution of Phenotypes across 4 Age Groups

Conclusions

In this study, we used HPO to annotate phenotypes from 2010-2015 clinical notes at Mayo Clinic. Phenotypic characteristics were collected for patients' demographic and *Wilson's Disease*.

Acknowledgements

This work was made possible by internal funding from Center for Individualized Medicine of Mayo Clinic and NCATS Biomedical Translator Award, OT3TR00201901.

References

- [1] Robinson, Peter N., et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics* 83.5 (2008): 610-615.

Address for correspondence

Hongfang Liu, Ph.D. Mayo Clinic

Harwick Building, 205 3rd Ave SW, Rochester, MN 55905

Liu.Hongfang@mayo.edu

Using a Simulation Modelling Approach to Manage Outpatient Department Waiting Time at the National Hospital of Sri Lanka

S.N.Silva^a, R.H. Hewapathirana^b, W.M.A.U. Jayatilleke^c

^aHead of Academic Affairs (International Institute of Health Sciences, Sri Lanka)

^bChief Executive Officer (Health Informatics Society of Sri Lanka)

^cSenior Lecturer (Post Graduate Institute of Medicine, University of Colombo, Sri Lanka)

Abstract

A simulation modeling approach was used to determine the optimal human resource solutions for the main functions of the OPD by observing 384 conveniently selected patients at the National Hospital of Sri Lanka. This work seems to be optimized by increasing the morning shift doctors, causing minimal disturbance and inconvenience to the administration and the workers.

Keywords:

Hospital Departments, Resource Allocation, Waiting Lists

Introduction

The Out Patient Department (OPD) at the National Hospital of Sri Lanka (NHSL) faces congestion due to overcrowding, leading to forming long queues and causing delays in patient care. The objectives were to identify an ideal resource allocation solution to reduce outpatient waiting time in Medical consultation, Phlebotomy and Dispensing processes at the NHSL OPD.

Methods

A process analysis was done using observations of the OPD process, and interviewing the key informants [1]. A times study to evaluate the OPD process was done observing 384 conveniently selected patients. The arrival and service times were calculated using the above data [2]. A Discrete Event Simulation using the ARENA software was carried out to model the current processes and to evaluate different scenarios that can reduce the delays. Each simulation was run 100 times.

Results

The most congested process in the OPD was the consultation and the most congested time was the weekday morning shift. The average inter-arrival time was 21 seconds at this time. The mean consultation time was 224 seconds and was considered a constant. For this process when the medical officers are increased from 12 to 15, the waiting time would reduce by 4 minutes and 30 seconds (66.1% reduction) and the patients in the OPD queue reduced by 25 (81.2%). Similarly when the doctors for the night shift was increased to 3, the waiting would reduce by 11.1minutes (77.6%) and will shorten the queue by 80%.

Table 1– Percentage improvements of the OPD process from the suggestions made for the morning and night shifts

	Morning : Changing from (12 Doctors, 1 Nurse, 3 Pharmacists) to (15 Doctors, 1 Nurse, 3 Pharmacists)	Night : Changing from (2 Doctors, 1 Nurse, 1 Pharmacist) to (3 Doctors, 1 Nurse, 1 Pharmacist)
Work in Progress	18.8	5.8
(Reduction)	31.5%	42%
Waiting Time	4.5min	11.1min
(Reduction)	66.1%	77.6%
Total time at the OPD	4.5min	11min
(Reduction)	23.4%	40.1%
Time spent in the	5.9min	12.8min
Consultation queue	75.6%	80.5%
(Reduction)		
Patients in the	25	7.2
Consultation queue	81.2%	80%
(Reduction)		
Average Utilization	2.99	-13.8
of a Doctor	3%	-14.2%

Conclusion

The process can be optimized by increasing the morning shift doctors, causing minimal disturbance and inconvenience to the administration and the workers..

Acknowledgements

The authors would like to acknowledge the support from the administration of the National Hospital of Sri Lanka.

References

- [1] Salleh MN, Yusof N, Haji Ali H. Using simulation to reduce out-patient waiting time: A case study at jitra health center.
- [2] Afrane S, Appah A. Queuing theory and the management of Waiting-time in Hospitals: The case of Anglo Gold Ashanti Hospital in Ghana. International Journal of Academic Research in Business and Social Sciences. 2014 Feb 1;4(2):34

Address for correspondence

Dr. S. Nishan Silva, snishansilva@gmail.com, +94718464567

Data Mining Applied to Analysis of Contraceptive Methods Among College Students

Priscyla Waleska Simões^{a,b,c,f}, Samuel Cesconetto^{a,c}, Eduardo Daminelli Dalló^{a,c}, Maria Marlene de Souza Pires^d, Eros Comunello^e, Felipe Borges Tomaz^a, Eduardo Pícolo Xavier^a, Pedro Antonio da Rosa Brunel Alves^a, Luciane Bisognin Ceretta^b, Sandra Aparecida Manenti^{a,c}

^a Grupo de Pesquisa em Tecnologia da Informação e Comunicação na Saúde, Universidade do Extremo Sul Catarinense, Brazil

^b Programa de Pós-Graduação em Saúde Coletiva (Mestrado Profissional), Universidade do Extremo Sul Catarinense, Brazil

^c Curso de Medicina, Universidade do Extremo Sul Catarinense, Brazil

^d Programa de Pós-Graduação em Ciências Médicas, Universidade Federal de Santa Catarina, Brazil

^e Vision Lab, Programa de Pós-Graduação em Computação Aplicada, Universidade do Vale do Itajaí, Brazil

^f Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas, Universidade Federal do ABC, Brazil.

Abstract

The aim of this study was to use the Data Mining to analyze the profile of the use of contraceptive methods in a university population. We used a database about sexuality performed on a university population in southern Brazil. The results obtained by the generated rules are largely in line with the literature and epidemiology worldwide, showing significant points of vulnerability in the university population. Validation measures of the study, as such, accuracy, sensitivity, specificity, and area under the ROC curve were higher or at least similar as compared to recent studies using the same methodology.

Keywords:

Contraception; Data Mining; Medical Informatics

Introduction

Data mining consists of applying algorithms to identify and analyze information in order to produce patterns or models [1]. In this context, the object of the study is sexuality, which is considered inherent to life and health and is expressed early in humans as one of the indices that measure the level of quality of life [2]. This paper presents the profile and experience of sexuality in a university population in southern Brazil generated using classification techniques.

Materials and Methods

This is an applied and technological study and was approved by the Research Ethics Committee on Human Beings. The database used for this study addressed the sexuality profile and the vulnerability regarding sexually transmitted diseases (STD) and syndrome immune deficiency acquired (SIDA) in a university population in southern Brazil. The classification of data was carried out using the J48 algorithm and the Waikato environment for knowledge analysis software (WEKA). Only the rules belonging in trees with at least 70% of the records correctly classified, and associations in 5% of the sample were used. For the evaluation of the knowledge generated, we used accuracy.

Results

After preprocessing, we generated 84 decision trees which have provided 3955 rules that were evaluated and selected

according to the relevance to the objectives of the study. The accuracy is 78%, ROC area is 0.76 and the Kappa coefficient is 0.53. Sensitivity and specificity is 80.36% and 71.46% respectively; 80.38% instances were classified correctly and 19.64% instances were classified incorrectly. Among the evaluated rules it is highlighted that when selected as a class, the “use of condoms as a contraceptive method”, it was found that the majority use condoms as contraceptive method, nor use sporadically as a form of prevention of sexually transmitted diseases, Did not use it in the last 5 sexual relations, not did they use the morning-after pill because they used oral or injectable contraceptives.

Conclusion

In conclusion, the results obtained by the generated rules are largely in line with the literature and epidemiology worldwide, showing significant points of vulnerability in the university population. Validation measures of the study, as such, accuracy, sensitivity, specificity, and area under the ROC curve were higher or at least similar as compared to recent studies using the same methodology.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP), Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC).

References

- [1] R. Bellazzi and B. Zupan, Predictive data mining in clinical medicine: current issues and guidelines, *Int J Med Inform* 77(2), (2008), 81-97.
- [2] M.L. Heilborn, Entre as tramas da sexualidade brasileira. *Rev Est Fem* 14(1), (2006), 43-59.
- [3] P.W.A. Simões, P.J. Martins, R.A. Casagrande, K. Madeira, M.C. de Mattos, S.A. Manenti and et al, Using a model of parallel distributed processing associated with data mining in the characterization of sexuality in a university population. *Stud Health Technol Inform* 2013; 192:1135.

Address for correspondence

Prof PhD Priscyla Waleska Simões – pritsimoes@gmail.com

Identifying Chemical-Disease Relationship in Biomedical Text Using a Multiple Kernel Learning-Boosting Method

Yueping Sun, Yu Zhang, Jiao Li

Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, China

Abstract

Chemical-induced disease relations (CID) are crucial in various biomedical tasks. In the CID task of BioCreative V, no classifiers with multiple kernels have been developed. In this study, a multiple kernel learning-boosting (MKLB) method is proposed. Different kernel functions according to different types of features were constructed and boosted, each of which were learned with multiple kernels. Our multiple kernel learning-boosting (MKLB) method achieved a F1 score of 0.5068 without incorporating knowledge bases.

Keywords:

Machine Learning, Artificial Intelligence, Data Mining

Introduction

Chemical-disease relations are crucial in various biomedical tasks, such as developing new drugs and predicting potential toxicity. Automatic chemical-disease relation detection from free text remains challenging, and a key challenge in the task lies in integrating multiple types of data for accurate extraction. Recently, a chemical-induced diseases relation (CID) task and related corpus was provided by BioCreative V. Reviews of the team's work shows that Support Vector Machines (SVM)-based classifiers are selected but no work incorporate multiple kernel learning into the classifiers.

Multiple kernel learning refers to a set of machine learning methods that use a predefined set of kernels and learn an optimal linear or non-linear combination of kernels as part of the algorithm. SPG-GMKL (Spectral Projected Gradient-Generalized Multiple Kernel Learning) developed by Jain et al [1] provides a state-of-the-art specialized optimization algorithm. In this work, SPG-GMKL is applied to different feature groups and a boosting algorithm is proposed in addition to the basic classifiers.

Methods

The training and the development data sets provided by BioCreative V are combined as a training set. The evaluation metrics include F1 score, precision and recall.

We systematically extracted the following features to train the classifiers, including: (1) Context words with position (-4,-3,-2,-1,1,2,3,4); (2) Context words with position in a co-occurring sentence; (3) Minimal middle string between chemical mentions and disease mentions; (4) Parser results of co-occurring sentences; (5) Binary features, including whether the chemical is head chemical, chemical in title, disease in title, etc. As shown in Figure 1, when the SPG-GMKL module was used for SVM implementations respectively for different features, the results of the basic classifiers are further boosted using algorithms such as linear function and voting function (MKLB).

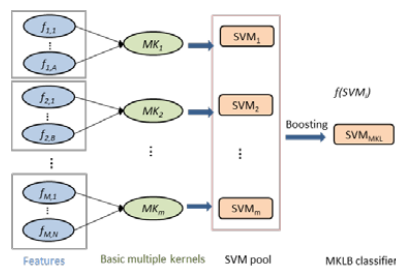


Figure 1– Illustration of the process of MKLB

Results

As shown in table 1, the best classifier is MKLB with voting as the boosting method, with F1 score of 0.5068, which is similar to the Entity plus Context score of CD-REST. [2]

Table 1– CID performance of MKs and MKLB

Method	P	R	F
SVM	0.6227	0.2880	0.3938
MK1	0.4074	0.3902	0.3987
MK2	0.5099	0.2655	0.3492
MK3	0.4969	0.1510	0.2317
MK4	0.5240	0.1126	0.1853
MK5	0.3952	0.3077	0.3653
MKLB	0.5486	0.4736	0.5068

Conclusion

We introduced a multiple kernel learning-boosting (MKLB) method to identify CID in biomedical text, and got a considerable F1 score without incorporating knowledge bases. This framework is extensible, making it more convenient to incorporate new features.

Acknowledgements

This study was supported by National Natural Science Foundation of China (Grant No. 81601573) and The National Key Research and Development Program of China (Grant No. 2016YFC0901901)

References

- [1] A. Jain, S. V. N. Vishwanathan and M. Varma. SPG-GMKL: Generalized multiple kernel learning with a million kernels, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2012:750-758.
- [2] J. Xu, Y. Wu, Y. Zhang, et al. CD-REST: a system for extracting chemical-induced disease relation in literature, Database, 2016:baw036.

Address for correspondence

Jiao Li, Ph. D., Email: li.jiao@imicams.ac.cn

Automated Report Generation for Research Data Repositories: From i2b2 to PDF

Volker S. Thiemann, Tingyan Xu, Rainer Röhrig, and Raphael W. Majeed

Department of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany

Abstract

We developed an automated toolchain to generate reports of i2b2 data. It is based on free open source software and runs on a Java Application Server. It is successfully used in an ED registry project. The solution is highly configurable and portable to other projects based on i2b2 or compatible factual data sources.

Keywords:

Databases, Factual; Medical Controlling; Emergency Service, Hospital

Introduction

The AKTIN [1] project is a German national emergency department (ED) registry initiative. In this project we use i2b2 databases to establish local data warehouses for emergency room (ER) data in participating hospitals. One part of the project is to provide automated data reports intended for quality reports, research feasibility analysis and data entry controlling. The key requirements are (1) automatic execution on Java Application Server, (2) application of standard formats and tools and (3) free open source software (FOSS).

Methods

The data source for this project is an i2b2 database. It is based on PostgreSQL and contains factual data in a star schema, common for clinical data warehouses.

The export is done via the open source tool HStream [2]. Based on a XML configuration file with details about the concept relations, HStream exports the i2b2 data in a tabular (CSV) format.

The statistical evaluation of the data is done via script with the R project statistical computing software. The outputs of that script are vector graphic plots (SVG) and XHTML tables.

The PDF is then created using Apache Formatting Objects Processor (FOP). The XSL-FO transformation is broken down into steps to facilitate maintenance: (1) XSL-FO template, (2) XML template with text structure and content, (3) Client configuration file (e.g. hospital name), (4) File reference list for plots and tables.

Results

The developed reporting toolchain meets all requirements and was released to our clinical partners. It is currently used for three different reports in the AKTIN project. The reports can be requested in a web interface or they can be scheduled to be periodically sent via email. The modular structure allows to

add reports by writing XML configuration and R scripts for plots and tables.

The execution time depends on the amount of data, the extent of the report and the hardware. It ranges from less than a second for simple reports to less than a minute for several pages and several thousand records.

Discussion

Within the boundaries of our requirements we could have used LaTeX as an alternative to produce the PDFs. We chose Apache FOP because it is less complex, has fewer dependencies and runs natively in a Java environment.

While there are other statistical software packages, the R Project was the only actively maintained free open source software. Additionally, it runs on all major operating systems and is very popular in the scientific community.

It can be argued that PDF is not the only possible output format. Yet, it is the cross-platform de-facto standard for archiveable ready-for-print documents. HTML was not favoured, because of its inferior print layout and its multiple files, which are unsuitable for emailing and archiving.

Conclusion

We were able to produce PDF reports from i2b2 data, using a generic, standards based approach. The reports are highly configurable and portable to other projects based on i2b2 or other clinical data sources.

Acknowledgements

Funding by BMBF No. 01KX1319B

References

- [1] J. Ahlbrandt, D. Brammen, R.W. Majeed, R. Lefering, S.C. Semler, S. Thun, F. Walcher, R. Röhrig, Balancing the need for big data and patient data privacy--an IT infrastructure for a decentralized emergency care research database, *Stud Health Technol Inform* (2014), 205:750-4.
- [2] R.W. Majeed, R. Röhrig, Automated realtime data import for the i2b2 clinical data warehouse: introducing the HL7 ETL cell, *Stud Health Technol Inform* (2012), 180:270-4.

Address for correspondence

Volker Sebastian Thiemann

Email: volker.thiemann@uni-oldenburg.de

Extraction and Evaluation of Medication Data from Electronic Dental Records

Yue Wang^a, Zasim Siddiqui^a, Anand Krishnan^{a,b}, Jay Patel^{a,b}, Thankam Thyvalikakath^{a,c}

^aDental Informatics Core, Indiana University School of Dentistry, Indianapolis, Indiana, USA

^bIndiana University School of Informatics and Computing, Indianapolis, Indiana, USA

^cRegenstrief Institute, Indianapolis, Indiana, USA

Abstract

With an increase in the geriatric population, dental care professionals are presented with older patients who are managing their comorbidities using multiple medications. In this study, we developed a system to extract medication information from electronic dental records (EDRs) and provided patient distribution by the number of medications.

Keywords:

Dental Informatics, Electronic Dental Record, Natural Language Processing

Introduction

Preoperative evaluation of dental patients' medication history is a vital component that can significantly affect the dental care provided. Medication information extraction using natural language processing (NLP) was the theme for the 2009 Informatics for Integrating Biology and the Bedside (i2b2) challenge[1]. Following in the footsteps of the i2b2 challenge, we evaluated the medication lexical variants that existed in our dataset and the patient distribution according to the number of medications.

Methods

Cohort Identification

Electronic dental records (EDRs) were extracted from Indiana University School of Dentistry (IUSD) for adult patients who went under comprehensive oral evaluation from January 1st, 2009 to December 31st, 2011. A total of 11,220 unique patients de-identified records were used for this study.

Medication-related Term Identification and Extraction

Three data fields from medical history form was used for this study: "Do you take any medications?", "What medications are you currently taking?" and "Have you taken any other medication in the past 5 years". The data were first tokenized to generate singular terms and any duplications of these terms were removed. Two reviewers independently identified the medication related term (terms containing drugs name: either brand name or generic name) and non-medication related term (any term except medications such as noun (patients), verbs (taking), etc.). Any disagreement between them were resolved through discussion. These medication-related terms was then ran against the 11,220 patient records.

Descriptive statistics were used for this study. Medication lexicon profile and patient distribution are presented.

Results

Table 1. Medication Lexicon Profile Generated by Reviewers
Medication names were an exact match to a drug name. Medication-related terms were lexical variants.

	N	Percentage
Medication names	1,022	18.35%
Medication-related Terms	4,547	81.65%
Total	5,569	100.00%

Table 2. Patient Distribution by Number of Medications.

Number of Medications	Male	Female	Total
0	2,725	2,564	5,289
1 to 3	1,456	1,854	3,310
4 to 6	531	825	1,356
7 to 9	262	284	546
10 to 12	126	197	323
13 to 15	62	97	159
16+	50	106	156
Total	5,212	5,927	11,220

Conclusion

This study demonstrated the need to capture lexical variants for medication extraction from EDRs. Additionally, this study showed that majority of dental patients are taking at least one medication.

Acknowledgements

We would like to thank Dr. Mei Song for her valuable comments on the manuscript.

References

- [1] O. Uzuner, I. Solti, and E. Cadag, Extracting medication information from clinical text, *J Am Med Inform Assoc* 17 (2010), 514-518.

Address for Correspondence

Thankam Paul Thyvalikakath, DMD, MDS, PhD
Associate Professor & Director of Dental Informatics Core
Research Scientist, Regenstrief Institute, Inc.
Indiana University School of Dentistry, IUPUI
1121 W Michigan Street, Room DS314
Indianapolis, IN 46202
Email: tpt@iu.edu Phone: (317) 274-5460

Development of Asian Non-Small Cell Lung Cancer Survival Prediction Model Using an Innovative Method of Bayesian Network

Yin Bincan, Xin Shichao, Zhao Yuhong

Department of Medical Informatics, China Medical University, Shenyang, Liaoning, China

Abstract

We constructed a novel prognostic model using an innovative method of Bayesian Network (BN) to predict Non-Small Cell Lung Cancer survival status within 5 years after operation in the Asian population. The proposed BN model could present the relationship between prognostic factors and showed the highest performance among other machine learning (ML) algorithms.

Keywords:

Bayes Theorem; Carcinoma, Non-Small-Cell Lung; Prognosis

Introduction

Lung cancer is the major cause of death for tumor patients, with approximately 83% of all lung cancer cases being Non-Small Cell Lung Cancer (NSCLC) and the overall 5-year survival rate is only 22.1% [1]. Physicians usually focus on scanty prognostic points but neglect other factors, which receive unsatisfactory prognosis prediction. We built this model in order to find the relationship between prognostic variables as well as predict the probability of patients' 5-year survival status to solve the gap in knowledge about the NSCLC prognostic system.

Methods

The cohort included Asian NSCLC patients whose follow-up period over 5 years (N=683) from SEER collected by the National Cancer Institute. Datasets were randomly split in training set (n₁=495) and test set (n₂=188) with a ratio of 3:1. The Holdout method was used in learning data, testing the performance of this model. We filtered 5 prognostic factors from original factors, which involved 18 independent variables by statistical analysis through SPSS. After data processing, to enhance the applicability of prognostic model, we used a new form method of BN, which is combined Tabu Search (TS) and clinical experience that distinguish from the traditional method, which only depend on data learning. TS is a heuristic algorithm based on neighborhood structures and iteration to solve optimization problems [2]. We applied TS to construct our model at first, then referred to physicians' advice to adjust the structure of the model by R. WEKA was for contrast experiment with other popular machine learning methods.

Results

STATUS is designated as "alive" if the patient can survive equal/longer than 60 months, or it is designated as "dead".

Five high relative factors were generated by logistic regression from training set: grade, tumor size, stage, age, lymph nodes ratio (LNR). LNR was the advice from physicians which is mixed with two other relative variables in this study. The predictive accuracy, F-Measure, and AUC was (72.87%, 0.719, 0.67) obviously exceeded three other ML methods: Decision Tree (67.02%, 0.654, 0.568), Support Vector Machine (68.62%, 0.662, 0.611) and Artificial Neural Network (64.89%, 0.589, 0.615). Five variables were selected to construct the ultimate predictive model (Figure 1).

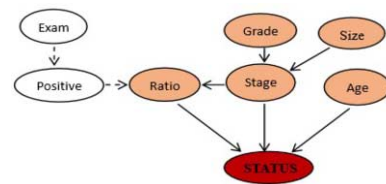


Figure 1-Asian NSCLC survival prognostic BN model

Conclusion

The most convincing procedure of modeling should be combined reliable datasets, proven methods, literature and experience from domain experts instead of merely applying conventional approaches. Compared with other network learning methods and main ML algorithms, the proposed BN prognosis predictive model showed outstanding performance, especially the accuracy, which could be considered guiding clinical decisions about the prognosis of neoplasm.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No.71473268) and the Social Development in Science and Technology Department of Liaoning Province (No.2013225079) to Dr. Zhao Yuhong.

References

- [1] National Cancer Institute. Cancer Statistics Review, 1975-2013 http://seer.cancer.gov/csr/1975_2013/sections.html
- [2] Glover F. Tabu Search—Part I[J]. Informs Journal on Computing, 1989, 1(1):89-98.

Address for correspondence

Zhao Yuhong, PhD, Department of Medical Informatics, China Medical University, email: joan@mail.cmu.edu.cn

Comparison of Different Algorithms for Sentiment Analysis: Psychological Stress Notes

Sunmoo Yoon^a, Faith Parsons^b, Kevin Sundquist^b, Jacob Julian^b, Joseph E. Schwartz^{b,c},
Matthew M. Burg^d, Karina W. Davidson^b, Keith M Diaz^b

^aSchool of Nursing, ^bCenter for Behavioral Cardiovascular Health, Columbia University Medical Center, New York, NY, USA;

^cDepartment of Psychiatry and Behavioral Science, Stony Brook University, Stony Brook, NY, USA;

^dDepartments of Internal Medicine and Anesthesiology, Yale University School of Medicine, New Haven, CT, USA

Abstract

To visualize and compare three text analysis algorithms of sentiment (AFINN, Bing, Syuzhet), applied to 1549 ecologically assessed self-report stress notes obtained by smartphone, in order to gain insights about stress measurement and management.

Keywords:

natural language processing

Introduction

Psychological stress is linked to all six of the most common causes of death in the U.S. In psychology, content analysis methods derived from paper-and-pencil surveys have been applied to patient records to improve mental health outcomes. With the advance of technology, there is an increasing volume of patient generated free-text data reporting mental health symptoms and context. As a result, natural language processing-computer linguistics has been successfully applied to patient-generated free-text to gain insights from symptom and emotion management. A sentiment analysis package, 'Syuzhet', for processing free-text data has recently become publicly available. However, few studies have applied this package to free-text stress notes or diaries extracted from smartphone-based ecological momentary assessments [1].

This study aims to visualize and compare three algorithms for sentiment analysis (Syuzhet, AFINN, Bing) applied to 1549 ecologically assessed self-report stress notes using smartphones to gain insights into how the analysis of large volumes of stress diaries might inform emotion management.

Methods

We extracted 1549 free-text notes describing self-reported momentary stressful occurrences, which were collected daily from Jan 2014 to April 2015 from sixty participants. Natural language processing was applied using three sentiment analysis algorithms (Syuzhet, AFINN, Bing) [1]. Pearson correlations were calculated between each algorithm and the participant's concurrently self-reported stress rating (0-10 scale).

Results

Figure 1 displays the pooled emotion scores from 1549 stress notes, each applying a different sentiment analysis. Pearson correlation coefficients among the three algorithms and self-rated stress scores are shown in Table 1. The correlations among the three algorithms are moderately high, but the correlations of algorithm scores with self-ratings are low. Positive emotion (lack of negative feeling) was detected from

half of the corpora of stress notes. (e.g., "Excitement!" Syuzhet emotion score +1, Self-report stress score -4).

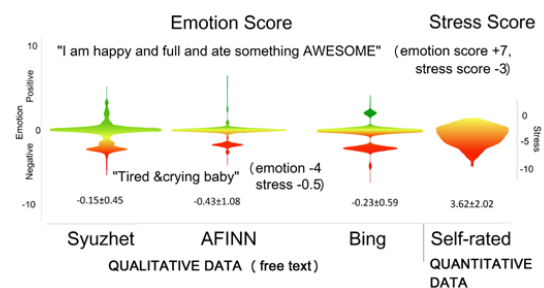


Figure 1 – Visualization of Distribution of Emotion Scores of Daily Stress Notes applying Different Algorithms

Table 1 – Correlations among Three Sentiment Algorithms

Algorithms	Syuzhet	AFINN	Bing
Syuzhet	1		
AFINN	0.73**	1	
Bing	0.83**	0.67**	1
Self-Report Score	0.04	0.03	0.03

**p< 0.01, N=1549 notes

Conclusion

Application of sentiment analysis natural language processing and visualization techniques provide insights for research teams regarding large volumes of daily self-report stress notes. The positive emotion scores detected by sentiment analysis algorithms from qualitative data (free text) provide quantified descriptive contextual information on low level self-rated stress scores.

Acknowledgements

National Institute of Health grant # R01 HL115941; NSF Institute for Pure and Applied Mathematics

References

- [1] M. Jockers, Package Syuzhet, (2016), V 1.0.

Address for correspondence

630W 168 street, mail code 6, New York, NY, 10032 USA,
sy2102@columbia.edu

The Study of Hot Spots on Hepatitis B Dissertation Based on Co-Word Analysis in China

Zhang W^a, Wang Y B^b, Zhang X Z^a, Duan H M^a

^aThe School of Nursing, Beijing University of Chinese Medicine, Beijing, China

^bPeking University of Software & Microelectronics, Beijing, China

Abstract

Doctoral dissertations and master's theses of "hepatitis B" (HB) in CNKI database were retrieved to explore the hot spots on HB research field in China, we processed, visualized the data above, and analyzed those data by using co-word analysis informetrics method. Then, the figure for co-occurrence analysis of high-frequency keywords in theses on HB were plotted, which represented eight topics, could help health staff to understand hot topics in this field.

Keywords: Hepatitis B; Co-word analysis; Dissertation

Introduction

HB is a kind of the most wide-spread and most serious harmful infectious disease in China, which is caused by hepatitis B virus (HBV), has got more and more attention by clinical researchers^[1]. In general, any doctoral dissertation and master's thesis is a complete and systematic presentation in a particular field, it is more profound and more professional than other type of literature of a particular field. Keywords are most reverent to the article which can reflect its core contents. So, to analyze the co-occurrence of high frequency keywords could represent hot spots in a particular research field^[2].

Methods

We collected data from digital dissertations and theses indexed by CNKI (China National Knowledge Infrastructure) database in China. We retrieved the subject term "hepatitis B" OR subject term "HBV" from 1980 to April 10th, 2016. As a result, a total of 7430 papers were found.

The data was processed and analyzed mainly by computer programs, SPSS software (version 18.0). Firstly, we calculated high-frequency keywords after the removal of abnormal data. During this step, in order to get more precise results, general words with no actual meaning were removed and the synonyms of keywords were merged. For example, "hepatocellular carcinoma" and "liver tumors" are replaced by "liver cancer". After processing the keywords, word-frequency was calculated by computer programs and 65 high-frequency keywords were identified. Secondly, a 65*65 co-occurrence matrix was constructed with computer programs. The value of the matrix represents the co-occurrence of a pair of keywords. Using Ochiai equivalent co-efficient method, dissimilarity matrix was constructed to calculate dissimilarity between a pair of keywords. Thirdly, hierarchical cluster analysis was conducted based on the dissimilarity matrix above with SPSS software, dendrogram was plotted by euclidean distance.

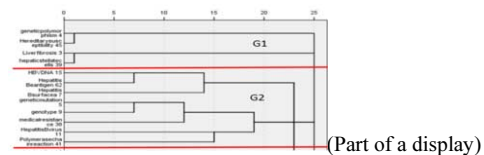


Figure1 - Results of HB high-frequency words by cluster analysis

Results

In total, 65 keywords were classified into eight categories according to their academic meanings, shown in Figure 1.

Conclusion

In general, we explore eight hot topics based on HB related dissertations and theses in CNKI database, which meet the contents in scientific research and clinical work of HB.

Acknowledgements

There is no declared of conflict in interest among all authors. The idea, design and revision of this paper was completed by Duan Hongmei, the primary writing of this paper was done by Zhang Wei, the methods part of the paper were done by Wang Yibo and all the data were collected by Zhang Xianzhen.

Reference

- [1] Chinese Society of Hepatology and Chinese Society of Infectious Diseases, Chinese Medical Association. The guideline of prevention and treatment for chronic hepatitis B: a 2015 update, Chinese Journal of experimental and clinical infectious diseases (Electronic Edition) **5** (2015), 570-589.
- [2] L. Feng and F.H. Leng, The development of theoretical studies of co-word analysis method, China Library Journal, **2** (2006), 88-92.

Address for correspondence

Corresponding author: Duan Hongmei;

Address: Beijing University of Chinese Medicine, Chaoyang District,
Beijing, China. Mail address: duanhongmei66@163.com

This page intentionally left blank

III. Human, Organizational, and Social Aspects

This page intentionally left blank

Nursing Informatics Pioneers Embrace Social Media

Juliana J. Brixey^a, Susan K. Newbold^b

^a School of Biomedical Informatics and School of Nursing, University of Texas Health Science Center, Houston, Texas, USA

^b Nursing Informatics Boot Camp, Franklin, Tennessee, USA

Abstract

The American Medical Informatics Association (AMIA) established the Nursing Informatics History Project to recognize the pioneers of nursing informatics. Fundamental to the pioneers was dissemination of knowledge. The purpose of this review was to identify pioneers who have embraced social media as of 2016. It is suggested that the pioneers participate in the advancement of nursing informatics via social media.

Keywords:

Nursing Informatics; History of Nursing; Social Media

Introduction

The AMIA Nursing Informatics History Project was established to record and maintain the history of nursing informatics [1]. As part of the History Project, 143 individuals were designated as pioneers of which thirty three were interviewed and videotaped. Pioneers were defined as people who were innovators, trailblazers, or groundbreakers in some aspect of nursing informatics, first to open a new area and prepare a way for other nurses to follow [2, 3]. The designated pioneers advanced nursing informatics through innovation in health information technology (HIT), education, administration, and research. Central to their work was dissemination so that others could utilize their findings and technology to advance nursing and nursing informatics. This study is an extension of a systematic review used to identify manuscripts published by the pioneers in PubMed for the years 2010-2015 as evidence of continued contributions to nursing informatics [4]. Therefore, the purpose of this project was to identify contributions to the field of nursing informatics as evidenced by participation in social media – specifically Twitter and LinkedIn. These applications were deemed mostly likely to be utilized for a public professional presence.

Methods

A systematic examination of Twitter and LinkedIn was conducted using each of the thirty three pioneer's first and last name to establish ownership of a social media account. Each Twitter account was analyzed to determine date joined, number of followers, number following, and the number of tweets sent. Each LinkedIn account was analyzed to establish the total number of connections and number of connections to other pioneers. Descriptive statistics were used to analyze the data

Results

Thirteen of the 33 pioneers interviewed were found to have Twitter accounts. The oldest account was established in 2008 with the most recent in 2015. The number of followers ranged

from 1 to 2994. Similarly the pioneer followed from 6 to 277. The number of tweets ranged from 0 to 2775.

A total of 28 pioneers were identified as having LinkedIn accounts with 21 fully accessible for analysis and 5 did not have a LinkedIn account. The average number of pioneer to other pioneers averaged at 9.5 with a range of 1 to 15.

Conclusions

Nurses use social media to share knowledge and ideas, debate issues, connect with communities, make announcements, and ask for information via professional social networking sites such as Twitter and LinkedIn.

The pioneers in nursing informatics continue to contribute to the dissemination of new knowledge in via social media. Moreover, the lack of use of social media by some pioneers could be attributed to the age of the pioneer, retirement, or lack of perceived value in participating in professional social networking. Future studies will include pioneers not interviewed to more fully quantify the use of social media for dissemination of knowledge.

References

- [1] AMIA. Nursing Informatics History Project [Website]; 2016. Available from <https://www.amia.org/working-groups/nursing-informatics/history-project>.
- [2] B. L. Westra and S. K. Newbold, American Medical Informatics Association Nursing Informatics History Committee, *CIN* March/April (2006), 113-115.
- [3] S. K. Newbold and B. L. Westra American Medical Informatics Association Nursing Informatics History Committee Update, *CIN* July/August (2009), 263-265.
- [4] Newbold, S.K. & Brixey, J.J. (2016). Nursing Informatics Pioneers Continue to Influence the Profession: A Sustainable Impact. Poster. In W. Sermeus, P. M. Procter, and P. Weber (Eds). *NI2016 proceedings: eHealth for All: Every Level Collaboration – From Project to Realization* (Vol. 225). Amsterdam: IOS Press.

Address for correspondence

Juliana.J.Brixey@uth.tmc.edu

Virtual Reality in Nursing: Nasogastric Tube Placement Training Simulator

Kup-Sze Choi

Centre for Smart Health, School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China

Abstract

A virtual reality (VR) based training simulator developed for nasogastric tube (NGT) placement is presented. It leverages the advantages of VR technology – safety, flexibility, interactivity and quantitative assessment – to speed up the learning curve of NGT placement. The simulator demonstrates the potential of VR for nurse education.

Keywords:

Computer Simulation, Nursing

Introduction

Virtual Reality (VR) technologies have been advancing in the past three decades, which proliferates a wide range of applications. In health care, representative examples include virtual surgical simulators developed for training, pre-operative planning and rehearsal [3]. VR also finds applications in nursing. This paper focus on the use of VR on the training of nasogastric tube (NGT) placement, a common clinical skill.

NGT placement concerns the insertion of a plastic tube through the nose into the stomach. Since the tube cannot be seen during insertion, it may be accidentally inserted into the lung, which can lead to complications or even fatality. Training of NGT placement is conventionally conducted by practising on humans, which causes discomfort, or on rubber manikins, which is static and does not replicate human anatomy and the insertion forces.

Method

To enhance the training, VR technology is employed to develop a computerized interactive training system. It is based on human anatomy and provides visual, audio and haptic feedback. A virtual patient (upper part of the body), a virtual NGT, and the virtual hand of the user are displayed on the screen. User can control the virtual hand using a 3D user interface, which is also a force feedback device that can produce real-time insertion forces. Verbal command can be used to instruct the patient to swallow (visualized by movement of Adam's apple) to facilitate the tube's advancement. The virtual patient can gag or cough (by playing audio clips) depending on the position that the tube is reaching. The insertion forces are modelled based on physics using finite element methods [1] or expert experience using fuzzy inference [2]. Quantitative metrics including the tube's location, speed and insertion force can be recorded.

Results

The VR-based NGT placement training simulator is shown in Figure 1. Nine clinical teachers were asked to comment on the system's the feasibility for nurse training and the overall real-

ism. A pilot study involving 80 nursing students was also conducted to evaluate system usability and their acceptance of the VR-based training approach. The feedback was positive in general, while suggestions to improve the user interface, to simulate more patient responses, and to extend the simulation for neonates were received.



Figure 1– The Training Simulator for NGT Placement

Conclusion

Taking advantages of the VR technology, the novel NGT placement training simulator provides an automated and standardized method that can increase the learning opportunity and enables asynchronous self-learning. It has the potential to speed up the learning curve so that students can be well-prepared for practice and operation in reality sooner.

Acknowledgements

The work is support in part by the Research Grants Council of the HKSAR (PolyU 5134/12E) and the Y.C. Yu Scholarship for the Centre for Smart Health.

References

- [1] K.-S. Choi, X. He, V.C.-L. Chiang, and Z. Deng, A virtual reality based simulator for learning nasogastric tube placement, *Computers in Biology and Medicine* **57** (2015), 103-115.
- [2] K.S. Choi, X.J. He, V.C.L. Chiang, Z. Deng, and J. Qin, A heuristic force model for haptic simulation of nasogastric tube insertion using fuzzy logic, *IEEE Transactions on Haptics* **9** (2016), 295-310.
- [3] D. Sorid and S.K. Moore, The virtual surgeon [virtual reality trainer], *IEEE Spectrum* **37** (2000), 26-31.

Address for correspondence

Corresponding author: Kup-Sze Choi
 Mailing address: Centre for Smart Health, School of Nursing
 The Hong Kong Polytechnic University
 Email: kschoi@ieee.org

Digital Health Information Workforce Planning Through a National Census

Kathleen Gray^a, Kerryn Butler-Henderson^b

^a Health and Biomedical Informatics Centre, University of Melbourne, Melbourne, VIC, Australia,

^b Australian Institute of Health Service Management, University of Tasmania, Launceston, TAS, Australia

Abstract

This poster describes the rationale and method for the development of a minimum data set as the first step toward an Australian national health information workforce census. Critical importance is attached to open and inclusive models, and the early and extensive engagement of key stakeholders, to provide a solid base for broader research into this workforce.

Keywords:

Censuses, Health Manpower

Introduction

Progress to advance digital health is occurring in health systems around the world, and a specialised digital health information workforce is integral to such change. Countries vary in their understanding of the nature and scope of the workforce concerned [1-4]. In Australia, peak health information profession bodies held national joint summit meetings in 2015 and 2016, to share longstanding concerns over skilled workforce shortages, uneven career structures and uncertain future configurations [5-6]. Our census project responds to the need flagged at the 2015 summit, to accurately identify Australia's current health information workforce, and lays a foundation for ongoing study of health information professions.

Method

A Health Information Workforce Census Management Group was formed, with representation from key associations and agencies - Australasian College of Health Informatics, Health Information Management Association of Australia, Health Informatics Society of Australia, Health Libraries Australia, State government health workforce officials, and the Australian Digital Health Agency – and also wider engagement with other interest groups. Whilst workforce censuses exist in a number of health fields, very little is published on the methodology used in each. After reviewing indirectly related work [7-12], we decided to use the WHO guidelines to develop a minimum data set, and to use a Delphi approach to gather expert input and consult widely about the data elements required.

Results

The development of the minimum data set for the health information workforce census is being conducted in six rounds of qualitative and quantitative data collection that started in February 2017. In round six, a survey instrument will be constructed based on the agreed minimum data set, and it will be uploaded into an online environment, where members of

the Management Group, Expert Panel and Consultation Group will pilot test it prior to its full deployment in early 2018.

Conclusion

This project appears to be the world's first aiming to use a national workforce census to generate empirical evidence about the health information professions broadly. It is a step toward improving knowledge about the specialised digital health expertise that is being developed in the health information professions, how this expertise is being deployed, and what effects it has on the safety and satisfaction of patients and clinicians, the sustainable management of health services, and the accountability of providers and funders of digital health initiatives. Given the globalisation that is facilitated by digital health, our research has implications for healthcare organisations and health systems internationally.

References

- [1] T. Eardley, *NHS Informatics Workforce Survey*. Assoc Informatics Professionals in Health & Social Care, 2006.
- [2] W. Hersh, The health information technology workforce, *Applied Clinical Informatics* 1 (2010), 197–212.
- [3] Canada Health Infoway, CHIMA, CIHI, & COACH, *Health Informatics and Health Information Management: Human Resources Outlook 2014–2019*, 2014.
- [4] X. Qi, Y. Wang, L. Xia, Y. Meng, Y. Li, et al., Cross-sectional survey on public health informatics workforce in China, *Public Health* 129 (2015), 1459–1464.
- [5] F.J. Martin Sanchez, & K. Gray, Recognition of health informatics in Australian standard classifications, *Stud Health Technol Inform* 204 (2014), 92–97.
- [6] K. Butler-Henderson, R. Lawrance, S. Low, N. Donnelly, & J. Lee, Researching the health information workforce, *Health Information Management Interchange* 6 (2016), 32–35.
- [7] WHO, *Human Resources for Health Information System: Minimum Data Set for Health Workforce Registry*, 2015.
- [8] W.T.F. Goossen, P.J.M.M. Epping, et al., Development of the Nursing Minimum Data Set for the Netherlands (NMDSN), *J Adv Nurs* 31 (2000), 536–547.
- [9] R. Ranegger, W.O. Hackl, & E. Ammenwerth, Development of the Austrian nursing minimum data set (NMDS-AT), *Stud Health Technol Inform* 212 (2015), 73–80.
- [10] J.E. Gibson, Using the Delphi technique to identify the content and context of nurses' continuing professional development needs, *J Clin Nurs* 5(1998), 451–458.
- [11] P. Mac Neela, P.A. Scott, M.P. Treacy, & A. Hyde, Nursing minimum data sets, *Nurs Inq* 13 (2006), 44–51.
- [12] J.P. Leider, et al., Methods behind PH WINS. *J Public Health Manag Pract* 21 Suppl 6 (2015). S28–35.

Address for correspondence: kgray@unimelb.edu.au

Development of a Data Collection Tool for MbHIS-QUAL: Evaluation of the Quality of Morbidity Data in Routine Health Information Systems (RHISs) in Hospitals

Lyn Hanmer^a, Edward Nicol^a, Debbie Bradshaw^a

^aSouth African Medical Research Council, Cape Town, South Africa

Abstract

The quality of morbidity data in multiple routine inpatient records in a sample of South African hospitals is being assessed in terms of data accuracy and completeness. Extensive modification of available data collection tools was required to make it possible to collect the required data for the study.

Keywords:

Health Information Systems, Data Accuracy, South Africa

Introduction

The evaluation of the quality of data in routine health information systems (RHISs) is an important component of assessing their potential to provide data to support planning, surveillance and patient care. The MbHIS-QUAL study is currently underway to assess the quality of morbidity data in routine inpatient records in South African public hospitals. While there have been multiple frameworks developed for collecting the actual data required for assessing the quality of data in RHISs, none of those identified was appropriate for this study. The closest model of the required data arose from guidelines for the structure and content of patient records [1], rather than from tools such as PRISM, which focus on the performance of RHISs at facility or higher organisational level [2].

Methods

The data collection tool for the MbHIS-QUAL study is being used to collect data from a sample of approximately 5780 routine patient records. The quality of the morbidity data in the records will be assessed in terms of accuracy and completeness. Data accuracy will be assessed at the hospital level by comparing information on patient functioning, procedures, and diagnoses recorded on paper-based systems (including routine patient medical records, discharge summaries, and ward registers) with the information captured in the electronic records for patients discharged during the study periods. Data completeness will be measured by assessing the proportion of discharge summaries that have all the required data fields completed by a clinician.

Results

Based on the guidelines for review of patient records defined by the Academy of Medical Royal Colleges [1] and other similar guidelines, the following data fields will form the basis of the review: patient ID, attending physician's signature, admission diagnosis, discharge date, discharge (final) diagnosis, condition on discharge, and procedures.

Even when using the guidelines for record reviews, further detailed work was required to determine actual data items to be extracted from the records, and some data items are still to be defined on the basis of actual data found in routine patient records. Some examples of issues addressed are listed below:

- To ensure patient confidentiality, the patient ID is replaced by a study ID assigned by the research team.
- The 'attending physician' could be a medical specialist, or a specialist in training, or a general practitioner, depending on the hospital and the level of staffing.
- There is currently no standard for recording 'condition on discharge' in patient records. A free text field has been defined to allow for recording any available data related to the condition of a patient which is present in the record.
- Data on diagnoses and procedures are not necessarily coded. Provision is therefore made for recording both free text and codes for these data elements.
- A maximum of three (3) procedures will be recorded for each study patient.

Conclusions

The conversion of published guidelines and tools to an appropriate and practical data collection tool for the MbHIS-QUAL study required significant effort. Further details of the data collection tool are available from the authors.

Acknowledgements

This work was funded by the SAMRC.

References

- [1] Health and Social Care Information Centre (HSCIC), Academy of Medical Royal Colleges. Standards for the clinical structure and content of patient records. London: HSCIC, 2013. Available from: <https://www.replondon.ac.uk/sites/default/files/standards-for-the-clinical-structure-and-content-of-patient-records.pdf>.
- [2] Aqil A, Lippeveld T, Moussa T, Barry A. Performance of Routine Information Systems Management (PRISM) Tools User Guide. Chapel Hill, (NC): MEASURE Evaluation, 2012. Available from: <https://www.measureevaluation.org/resources/publications/ms-12-51>.

Address for correspondence:

L.A. Hanmer lyn.hanmer@mrc.ac.za.

Digital Divide and Health Disparities in China: Data from a National Longitudinal Survey of CHARLS

Y. Alicia Hong^a, Zi Zhou^b, Ya Fang^b

^a School of Public Health, Texas A&M University, College Station, Texas, USA

^b School of Public Health, Xiamen University, Xiamen, China

Abstract

The China Health and Retirement Longitudinal Study (CHARLS, 2013) data was used to investigate internet use and mobile phone ownership in older Chinese adults and examine digital divide and social economic status and mobile technology adoption and health outcomes associations. Results suggest a significant digital divide associated with not only individual characteristics, but also neighborhood resources. Future eHealth programs should consider the accessibility of mobile tools and develop culturally appropriate programs for different social groups.

Keywords:

Internet; Cell Phones, Digital Divide.

Introduction

A large percent of the global population has access to the internet; mobile phone ownership is almost ubiquitous. A digital divide exists in almost every society and is more prominent in developing countries. China tops the world in terms of population, internet users, and mobile phone users, but data on its digital divide are rather limited. This study investigates internet use and mobile phone ownership in older Chinese adults (≥ 45 years), a population typically known for slow in adoption of technologies. We also examine the relationship between the digital divide and social economic status, and the association of mobile technology adoption and health outcomes.

Methods

We drew data from China Health and Retirement Longitudinal Study (CHARLS), a nationally representative survey with a sample size of more than 18,000. CHARLS household data of 2011 and personal data of 2013 were linked to capture information at the individual and community level. Multi-level logistic regression models were employed for data analysis.

Results

We found that among Chinese adults older than 45 years, only 4.1% had used the internet in the past month, 23% household had broadband connection at home, and 82% household owned mobile phones. Our multivariate logistic regression models showed that internet use was strongly associated with one's social-economic status (SES), rural-urban residence, neighborhood amenities, and community resources. Mobile phone ownership was strongly associated with SES and rural-

urban residence, but not so much with neighborhood amenities and community resources. Internet use was a significant predictor of self-reported health status and having disability, even after controlling for potential confounders at the individual, neighborhood, and community level. By contrast, mobile phone ownership was not significantly associated with health status or having disability.

Discussion

This study is one of the first to examine the digital divide and its relationship with health disparities in China. It provides important data for developing interventions to bridge the digital divide. It also suggests that future eHealth programs need to consider different the accessibility of mobile tools and develop culturally appropriate programs for different social groups.

Conclusion

Our study suggests that despite a big population of internet users and mobile phone owners, a significant digital divide exists in China, especially in Chinese older adults. Internet access is still limited to people with higher SES, but mobile phones are adopted by the general population.

Address for correspondence

Y. Alicia Hong, Ph.D. Associate Professor,
School of Public Health, Texas A&M University
1266 TAMU
College Station, TX, 77843, USA
Phone: 979-436-9343
Email: yhong@tamhsc.edu

Patient Handovers – Cognitively Demanding: Does the Handover EHR Meet This Challenge?

Ursula Hübner, Mareike Przysucha

Health Informatics Research Group, University AS Osnabrück, Osnabrück, Germany

Abstract

Patient handovers are crucial to warrant continuity of care and patient safety. The handoverEHR is an instrument that allows users to depict the main features of a clinical case in cognitive maps. We were interested whether this tool had an effect on the task load. A cross-over study with 30 nursing students was therefore conducted. Mental demand showed a statistical trend to be lower in cognitive map handovers than in the other handover types.

Keywords:

Patient Handoff; Electronic Health Record; Psychology, Medical

Introduction

Patient handovers between shifts of care givers are a well defined mechanism to warrant information continuity between care teams and prevent adverse events. In addition to the safety challenge, there is a time challenge which poses an emotional and mental demand on the persons involved [1]. The main research questions of this study therefore was, whether the use of an electronic tool, the handoverEHR [2], affected the subjective task load.

Methods

In order to answer the research question, a randomised controlled cross-over design with the three experimental tasks was chosen: a) handovers without the help of any instruments (WITHOUT=control group), b) handovers with handoverEHR and list presentations (LIST) and c) handovers with handoverEHR and cognitive maps (MAP). Thirty young nursing students in their third study year (23 females/7 males, 21.9±2.9 years old) handed over four patients in groups of up to four study participants during each experimental condition. Following these tasks, they were asked to fill in the NASA Task Load Index Questionnaire (NASA TLX).

Results

The results of the NASA TLX dimensions "mental demand", "temporal demand", "effort" and "frustration" are shown in Fig. 1. "Mental demand" decreased from WITHOUT, over LIST to MAP (62.8±16.7; 59.0±16.2; 55.4±21.2). This effect showed a trend for significance (F=2.6, df=2;58, p=0.08). Despite decreasing means in "effort" from WITHOUT to MAP (29.4±14.8; 25.1±12.4; 23.8±12.1), this effect was not significant. The lowest values were obtained for "frustration", however with a significant difference (F=4.03, df=2;58, p=0.04)

between the tasks: WITHOUT (13.3±15.0), LIST (12.0±10.9) and MAP (25.9±30.6). "Temporal demand" nearly stayed the same in all groups: WITHOUT (31.7±19.2), LIST (32.3±14.5) and MAP (33.7±17.8).

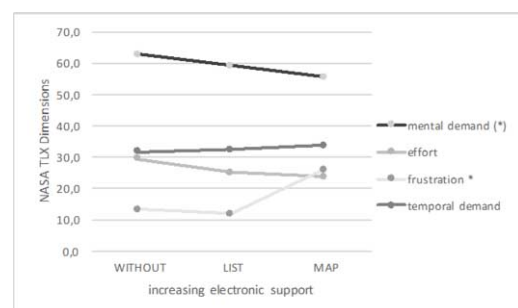


Figure 1: Four NASA TLX dimensions ((*) trend, * p<0.05)

Conclusions

We could demonstrate a trend for a positive effect of the handoverEHR on "mental demand". "Effort" and "temporal demand" were unchanged, while "frustration" increased significantly. These findings will be contrasted by results on the cognitive functions memory and decision making, which had been tested in the same research design.

Acknowledgements

This study was supported by a research grant (ZN2819) of the State of Lower Saxony, Hannover Germany.

References

- [1] J.Q. Young, R.M. Wachter, O. ten Cate, P.S. O'Sullivan, and D.M. Irby, Advancing the next generation of handover research and practice with cognitive load, theory *BMJ quality & safe* **25** (2016), 66–70
- [2] D. Flemming, M. Przysucha, and U. Hübner, Cognitive Maps to Visualise Clinical Cases in Handovers. Design, Implementation, Usability, and Attractiveness Testing, *Methods of information in medicine* **54** (2015), 412–23

Address for correspondence

Prof. Dr. Ursula Hübner Hochschule Osnabrück Health Informatics Research Group, D-49009 Osnabrück, Germany u.huebner@hs-osnabrueck.de, www.wiso.hs-osnabrueck.de/igw_team.html.

Evaluation of the Anonymity and Utility of De-Identified Clinical Data Based on Japanese Anonymization Criteria

Eizen Kimura^a, Satoshi Hasegawa^b, Koji Chida^b, Shoko Gamo^a,
Satoshi Irino^c, Haku Ishida^d, Yukio Kurihara^e

^a Department of Medical Informatics, Ehime University Graduate School of Medicine, Touon, Ehime, Japan

^b NTT Secure Platform Laboratories, Musashino, Tokyo, Japan

^c Department of Nursing, Ehime Pref. University of Health Sciences, Tobe, Ehime, Japan

^d Department of Medical Informatics & Decision Sciences, Yamaguchi University, Ube, Yamaguchi, Japan

^e Division of Health Informatics, Medical School, Kochi University, Kyoto, Kyoto, Japan

Abstract

We analyze the deterioration of clinical data quality due to anonymization. The result shows that data quality remained high with micro-aggregation and also verify the availability of noise addition to prevent illegal re-identification by matching another personal data.

Keywords:

Confidentiality, Anonymization, Privacy Preserving Data

Introduction

The Amended Act on the Protection of Personal Information in Japan, scheduled to go into full effect in the May of 2017, will open a way to provide medical records to third parties by de-identifying personal data. For de-identifying data, the law mandates the processing of personal information in accordance with standards stipulated by the rules of the Personal Information Protection Committee (PPC) [1] and the relevant guidelines. Currently, the PPC states that the five requirements must be met for the provision of data to a third party without personal agreement. At present, however, it is not clear what process will be sufficient to fulfill these rules because of the lack of technical guidelines for the anonymization process. The nature of an appropriate anonymization process may change depending on the content of the data and type of analysis. Our study examines use cases for clinical data to establish the best practice in anonymization.

Methods

We collected 7,200 quasi-healthy individuals with age, sex, and laboratory items: Hb, CR, ALT, RBC, WBC, aspartate aminotransferase, total protein, blood urea nitrogen, total cholesterol, and gamma glutamyl transpeptidase (GGT). Using the data, we examine various statistics, mean, standard deviation, quartile, the Mann-Whitney U-test (MU) and the Kolmogorov-Smirnov (KS) test for each group of the same age class and sex as a case study. Based on the age hierarchy required by researchers, ages were converted into four classes using 5-year intervals: 50–54, 55–59, 60–64, and 65–69 years.

As a requirement of the five rules, we evaluate k-anonymity assuming age and sex are quasi-identifiers. We also assumed that the exceptional description of a single attribute falls within three standard deviations (SDs) of all records and is within top and bottom 2.5% data within each same-sex group 5-year interval in age. We applied Maximum Distance to Average

Vector (MDAV[2]) for micro-aggregation. To reduce the possibility of identification by collating medical records with other data, we used Pk-anonymization[3] with noise addition and the Post Randomization Method (PRAM[4]).

Results

All data satisfied k=10 in the age and sex pairs. Therefore, k-anonymization processing was unnecessary for age and sex. We confirmed the errors were exceedingly small even if micro-aggregation withing 3 SDs is applied. The errors of RBC are smaller than those of GGT, the de-identified data with noise-addition cannot be used in either the basic statistics or hypothesis tests under existing conditions. The lower error rate of RBC compared with GGT may be attributable the wide range of GGT values, compared to the RBC values. Also, RBC has a closer distribution to the normal distribution.

Conclusion

We instantiated de-identified clinical data considering the requirements of Japanese anonymity criteria and verified the availability. Anonymization using only micro-aggregation maintained good data quality. However, we found that many sample data are required to obtain the statistical result with high quality for de-identified data with noise addition based on Pk-anonymity. In a future study, we will re-evaluate our results after improving the noise-addition algorithm and increasing the target data size, assuming large data processing.

References

- [1] Personal Information Protection Commission JAPAN, <https://www.ppc.go.jp/en/legal/> (2017).
- [2] J. Domingo-Ferrer and V. Torra, Ordinal, continuous and heterogeneous k-anonymity through microaggregation, *Data Mining and Knowledge Discovery* **11** (2005), 195–212.
- [3] D. Ikarashi, R. Kikuchi, K. Chida, and K. Takahashi, k-Anonymous Microdata Release via Post Randomisation Method, in: *International Workshop on Security*, Springer, 2015, pp. 225–241.
- [4] P. Kooiman, L.C.R.J. Willenborg, and J.M. Gouweleeuw, *PRAM: a Method for Disclosure Limitation of Microdata*, CBS, 1997.

Address for correspondence

Eizen Kimura, ekimura@m.ehime-u.ac.jp

Department of Medical Informatics, Ehime University Graduate School of Medicine

Developing an Evidence-Based Web-Education Program Designed to Help Nurses Learn to Care for Acutely Ill Intoxicated Patients in Emergency Departments

Marita Koivunen^{a,b}, Eija Noppari^c, Tiina Hakala^{a,d}, Anna-Kaisa Kautto^a

^a Satakunta Hospital District, Pori, Finland

^b Department of Nursing Science, University of Turku, Turku, Finland

^c Diaconia University of Applied Sciences, Pori, Finland

^d School of Health Sciences, University of Tampere, Tampere, Finland

Abstract

A significant part of the patients treated in an emergency department are under the influence of alcohol or drugs. It is important to identify individuals of patient groups who might benefit from an intervention to reduce drug use. The aim of the study was to develop an evidence-based web-education program aimed at enhancing nurses' knowledge and skills in the care, assessment, and management of substance-related disorders among acutely ill patients in emergency departments.

Keywords:

Emergency Treatment, Substance-Related Disorders, Computer-Assisted Instruction

Introduction

It has been estimated that in emergency departments about one third of the patients are under the influence of alcohol or drugs or are alcohol abusers [1]. The hazardous use of alcohol is a major global factor contributing to disease, injury, and death [2]. It is therefore important to identify individuals or patient groups who might benefit from an intervention in emergency departments to reduce alcohol use [3]. The care of acutely ill or injured patients who are also intoxicated requires wide-ranging competence among nurses working in emergency departments [4].

The aim of the study was to develop an evidence-based web-education program aimed at enhancing nurses' knowledge and skills in the care, assessment, and management of substance-related disorders among acutely ill patients in emergency departments.

Methods

At first, nurses' competencies to care for intoxicated patients in emergency departments were clarified using a quantitative survey (N=1220, n=252) in twenty emergency departments in Finland. Second, according to findings of the survey the fields of knowledge, which should take account in the web-education program, were formulated. We conducted preliminary evaluation of the content and learning methods of the web-education program with four nurses.

Results

The contents of the program were designed using results of the survey. The web-education program consisted of three parts: 1) patient centered encounter of intoxicated patients in

emergency care; 2) practical nursing methods in the care for intoxicated patients in emergency departments; and 3) developing quality of care for intoxicated patients in emergency departments. The preliminary evaluation showed that the program is easy to use and supports the nurses' knowledge.

Discussion

In practice, the nurses can complete the education program independently, but they have to get a feedback from the specialists. The nurses can seek information of the topic from the program. In addition, they can discuss on the web and use a diary, for example ethical questions that they were confronted with when they have cared for intoxicated patients.

Conclusion

The specific web-education program may be beneficial for nurses who care for the acutely ill intoxicated patients in emergency departments.

References

- [1] E. Bakke, S.T. Bogstrand, P.T. Normann, O. Ekeberg and L. Bachs, Influence of alcohol and other substances of abuse at the time of injury among patients in a Norwegian emergency department. *BMC Emerg Med* **16** (2016), doi: 10.1186/s12873-016-0085-2
- [2] A. Hankin, M. Daugherty, A. Bethea and L. Haley, The emergency department as a prevention site: A demographic analysis of substance use among ED patients. *Drug Alcohol Depend* **130** (2013), 230-233.
- [3] A. Trinks, K. Festin, P. Bendtsen and P. Nilsen, What makes emergency department patients reduce their alcohol consumption? - a computer-based intervention study in Sweden. *Int Emerg Nurs* **21** (2013), 3-9.
- [4] D. Mellor, M. McCabe, L.A. Ricciardelli, S. Brumby, A. Head, C. Mercer-Grant and A. Kennedy A (2012) Evaluation of an alcohol intervention training program for nurses in rural Australia. *J Res Nurs* **18** (2012), 561-575.

Address for correspondence

Marita Koivunen, email mhkoivu@utu.fi
phone +358447077887

Privacy Policy Implementation on the Nation-Wide EHR in Japan for Hospitals and Patients

Naoto Kume^a, Shinji Kobayashi^a, Kenji Araki^b, Hiroyuki Yoshihara^a

^a Department of Electronic Health Records, Kyoto University, Kyoto-city, Kyoto, Japan,

^b Department of Medical Informatics, University of Miyazaki Hospital, Miyazaki-city, Miyazaki, Japan

Abstract

Shared clinical information is an important contribution to regional medicine. Clinical information sharing with patients is also recommended to motivate patients and promote health. On the other hand, the threat of information leaks, caused by internet connected records, is critical to hospitals. The traditional approach is complete isolation of hospital networks, instead of information sharing. The authors propose methods here to maximize information sharing by following hospital preferences for electronic health records.

Keywords:

Access to Information, Electronic Health Records, Patient Participation

Introduction

Sharing information among hospitals is necessary to achieve collaborative medicine. Also, sharing information with patients will be mandatory to prevent serious illness, especially in chronic diseases.

However, information sharing is also a risk from the viewpoint of hospitals. Information leaks and the unnecessary sharing of information must be prevented. For example, information related to clinical activities, and not clinical results, does not need to be shared. Therefore, in Japan hospitals have traditionally hesitated to share information or have tried to minimize sharing as much as possible.

The establishment of a nation-wide electronic health record (EHR) system in Japan must tackle the problem of sharing or not sharing between hospitals and patients through policy. Therefore, the authors propose a structural access control that reflects the decisions of each stakeholder by prioritizing their privacy policies and preferences.

Methods

The authors designed a nation-wide EHR as a centralized system. All clinical information of the partner hospitals is stored in the centralized database. Firstly, the priorities of the privacy policy are defined in the following order: hospital, patient, and collaborative hospital. Secondly, a privacy policy matrix is provided to the information-sharing hospitals. Thirdly, patients are provided an opportunity to declare their preferences using a user interface portal to their own medical records.

Privacy policy matrix for information-sharing hospitals

A hospital declares their policy on a clinical item and department matrix. Clinical items are provided by the document classification definition from Medical Markup Language (MML) [1]. MML 4.1.2 provides seventeen items. The number of departments depends on the hospital. A hospital provides a decision to each item as open or closed to patients and collaborative hospitals.

Privacy policy for patients

A patient can access the list of hospitals, which is on his/her clinical history. A patient declares which hospitals can provide their information to other hospitals. For instance, if a patient has one record that should not be shared with another hospital, the patient can stop sharing the record, even if the hospital allows sharing with collaborative hospitals.

Results

The methods are implemented on an EHR, which connects forty hospitals. Clinical information, including sharing policies, are uploaded from each hospital. A feasibility study of administration operation is underway.

Conclusion

Prioritized policy making structure with a simple privacy policy matrix are achieved so that the proposed methods provide convincing grounds to hospitals agree to share maxim items.

Acknowledgements

This research is supported by “Clinical Study Oriented ICT Infrastructure Development Project - Sustainable Massive Health and Clinical Data Repository for Secondary Use” from Japan Agency for Medical Research and Development (AMED).

References

- [1] Medical eXtensible Markup Language Consortium, <http://medxmi.net/> (last access 2016.12)

Address for Correspondence

Naoto Kume, kume@kuhp.kyoto-u.ac.jp, KRP-9-506, Awatacho 91, Chudoji, Shimogyo-ku, Kyoto-city, Kyoto, JAPAN, 600-8815.

Patient's Adherence Level Determination System

Taranik Maksim^a, Georgy Kopanitsa^{a,b}

^a Tomsk Polytechnic University, Tomsk, Russia,

^b Tomsk State University for Architecture and Building, Tomsk, Russia

Abstract

The present paper deals with intellectual systems for patient's level adherence determination. The most popular nowadays technique for adherence evaluation is interview. We advise new automation approach, which compares direct patient data with treatment plan using fuzzy logic and standard ISO 13606. For this research data entered by users. The results can be useful for medical organisation with long-term patients.

Keywords:

Adherence, Informatics

Introduction

The quality of healthcare process and the results depend on the level of patient's adherence. World Health Organization (WHO) mentions, that increasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatment [1]. The question of adherence is the most relevant for patient with chronic diseases, cancer, stroke etc. Among the methods of adherence level determination there are implicit: interviews, question lists and explicit: real counters (as pill counter). Both methods have human factor, which provides uncertainty. Therefore, medical specialist cannot be sure in the accuracy of patient's answers. A great deal of modern adherence-based researches. However, the questions of automation of medical data acquisition and adherence level processing allowance of human factor, receive insufficient attention. Therefore, development of special intellectual automate systems for patient monitoring and adherence level evaluation is relevant modern issue. The presented paper deals with presentation, realization and evaluation of patient's level determination concept based on the direct data in the scope of long-term healthcare in cardiovascular department.

Methods

The intellectual system is based on the following concept. The treatment plan is included in the system by a medical professional. The system determines and formalizes some knowledge on injected medical drugs, the right amount of sleep, physical activity, specific diet that influence the patient's status. The system contains the objective data on patient's status based on the data obtained through ongoing monitoring portable devices. The junior medical staff includes this information in the system Enumerated list item. Based on the included data the system determines if the information matches the memorized treatment plan of the patient and assesses the treatment retention using fuzzy Mamdani algorithm. The obtained results are memorized. The duration of the present study on the treatment retention assessment was two weeks. We acquired the data of 20 patients

at the age of 20-45 (13 men, 7 women) that have a hypertensive disease diagnosis. The study was in hospital environment of the Scientific Research Institute of Cardiology (Tomsk) in the arterial hypertension unit.

Results

The first step of intellectual system approbation was treatment plan determination for all patients. The system collected patient's medical data and compared with treatment plan three times a day. As a result it was an evaluation (0-1). The mean of three evaluations was a day evaluation. After two weeks we calculated mean evaluation during the whole period (Table 1).

Table 1 – Results of patient's adherence level after 2 weeks

	Total evaluation	Patient state
Patient 1	0,5	improvement
Patient 2	0,5	improvement
Patient 3	0,5	improvement
Patient 4	0,4	inconspicuous
Patient 5	0,5	inconspicuous
Patient 6	0,5	inconspicuous
Patient 7	0,5	improvement
Patient 8	0,6	improvement
Patient 9	0,5	improvement
Patient 10	0,4	inconspicuous

We formalized patient's states, adopted to scale from 0 to 1, and got further results: recovery (0,75-1), improvement (0,5-0,75), inconspicuous (0,35-0,5), deterioration (less than 0,3). Based on the presented results it is possible to conclude, that developed system determined accurate adherence level of researched set of patients.

Conclusion

Presented in the current paper research is a first iteration of the work concern of development of adherence level intellectual instruments with further decision support process. These results allow to increase the accuracy and quality of adherence level determination.

References

- [1] A. Grover, H.S. Rehan, L.K. Gupta, M. Yadav. Correlation of compliance to statin therapy with lipid profile and serum HMGCoA reductase levels in dyslipidemic patients, Indian heart journal, (2016).

Address for correspondence

Georgy.kopanitsa@gmail.com (Georgy Kopanitsa)

Report and Analysis of Web Portal oldagesolutions.org for Indian Old Age Persons

Sushil Kumar Meher^a, A.B. Dey^b

^a Department of Computer Facility, All India Institute of Medical Sciences (AIIMS), New Delhi, India,

^b Department of Geriatric Medicine, All India Institute of Medical Sciences (AIIMS), New Delhi, India

Abstract

Oldagesolutions.org is a web portal for older persons in india created as apart of Technology Interventions for Elderly (TIE) intiated by the science and society division for benefit of elderly people keeping in view of the National Policy on Older People (NPOP). This is a collative effort of the multidisplinary team at All India Institute of Medical Sciences (AIIMS) comprising of doctors, genetic nurses, netritonist and physiotherapists who provide their valuable collective insights for an elderly centric approach.

Keywords:

Technology, National Health Programs, Aged

Introduction

India is home to 110 million older persons with high burden of diseases and disabilities and great need for access to affordable health care. The population of senior citizens in India is increased due to 60% increase in life expectancy in the last 60 years. Today India has the 2nd largest number of Elderly and every 7th senior citizen in the world is from India. There is a sharp increase in population of young-old 60 yrs to 69 yrs and old-old 80 yrs and above.

Oldagesolutions.org is a web portal for older persons in India created as a part of [Technology Interventions for Elderly \(TIE\)](#) initiated by the science & society division for benefit of elderly people keeping in view of the National Policy on Older People (NPOP). The web portal includes information related to health care and allied areas such as:

Recreation & Entertainment: indoor activities for seniors, outdoor activities for seniors, recreation centers

Physical Health: health and fitness, healthy eating, healthy bone, good sleep, taking care of your oral cavity, better sight in old age, immunization.

Mental Health: depression, living alone, worry & anxiety, cognitive impairment, bereavement.

Nutrition: nutritional considerations, nutritional requirements, dietary guidelines for life style, potential shortfall and over consumed nutrients, methods of assessment nutritional status.

Assistive Devices: occupational therapy, self care, mobility, house work, safety and security, leisure activities.

Facility: schemes & facilities, legal issues, safety of senior citizens a challenge. old age home directory, old age pension

Design & Environment: design & environment for elderly, barrier free environment

Methods

We propose 4-dimensional criteria to evaluate this website. The proposed criteria using a Likert scale are *content quality*, *design quality*, *organization quality*, and *user-friendly quality*. We have administered questionnaires to 272 participants.

Results

Analysis of the results of the 272 questionnaires is as follows:

1. *Content quality:* 48%- fully satisfied, 46% - satisfied, 5%- partial satisfied, 1% -not satisfied
2. *Design quality:* 42%- fully satisfied, 49% - satisfied, 07%- partial satisfied, 2% -not satisfied
3. *Organization quality:* 42%- fully satisfied, 46% - satisfy, 10%- partial satisfied, 2% -not satisfied
4. *User-friendly quality:* 40%- fully satisfied, 47% - satisfied, 12%- partial satisfied, 1% -not satisfied

Discussion and Conclusions

As per the above study, this web site provides a holistic service to ensure physical, mental, and social well being of senior citizens. More than 80% users are very satisfied with its content, design, and user friendliness.

Acknowledgements

Funding sources for the work is provided by Department of Science and Technology, Govt. Of India.

References

- [1] A. Abanumy, A. Al-Badi, and P. Mayhew, e-Government Website accessibility: in-depth evaluation of Saudi Arabia and Oman, *The Electronic Journal of e-Government* 3 (2005), 99-106.
- [2] American Nurses Association, *Nursing Informatics: Scope and Standards of Practice, 2nd Edition*, American Nurses Association, Silver Spring, MD, 2014.
- [3] A.K. Jha, D. Doolan, D. Grandt, T. Scott, and D.W. Bates, The use of health information technology in seven nations, *Int J Med Inform* 77 (2008), 848-854.

Address for correspondence

Dr. Sushil Kumar Meher, Department of Computer Facility
All India Institute of Medical Sciences, Ansari Nagar, New Delhi,
India-110029 ; Email: s_meher@hotmail.com
Phone : 91-11-26588332, Fax : 91-11-26588663
Cell : 9868397023

DOctimer: A Timing and Event Recording Tool for Direct Observational Research

Matthew E. Nolan^a, Brian W. Pickering^b, Vitaly Herasevich^b

^a Division of Pulmonary & Critical Care, Mayo Clinic, Rochester, MN, USA

^b Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, USA

Abstract

Clinical research often requires direct observation of clinicians performing routine tasks, but few effective data collection instruments exist. We describe the development of DOctimer - a web-based, platform-independent timing and counting application that allows researchers to easily record numerous tracking elements on a single screen, facilitating robust data collection for direct observational research.

Keywords:

Software; Data Collection; Medical Informatics Applications.

Introduction

The problem arose from our informatics research group's study needing direct observation of physicians using electronic medical record information during new admissions. We required a tool that could log the duration and/or count for more than 80 data elements, while allowing the observer to remain mobile and unobtrusive to the physician's workflow.

Several behavior timing/tracking solutions exist [1-3] but have shortcomings: lack of concurrently-running timer support, lack of task hierarchy support (i.e., group/subgroups), inability to log event sequences, inadequate interface for numerous items, lack of mobile-device compatibility, need for internet connection, or high cost. Therefore, we developed an easily configurable Direct Observation Counting timer (DOctimer) instrument to solve this research problem.

Methods

We developed a time tracking/counting/sequencing tool using HTML5, CSS and pure javascript as a single ".html" file that runs on any modern browser, including tablets, without needing an internet connection. The researcher inputs the data elements of interest using a highly structured nested object template that defines the data element name, type, and display style, with the template sequence and hierarchy used to imply groupings and generate html code. There are two core element types: 'timers' record all start/stop datetimes, total timing duration, and elapsed time, while 'counters' record the count and datetime events.

Results

Using an efficient single-screen interface, our tool can display over 80 timing and counting elements that are quickly recorded with a single "click." Hierarchical timing logic ('child' elements start their 'parent' elements) and mutual exclusivity (running timers stop when unrelated groups start) minimize user input requirements and improve recording speed. Results

are output in CSV or JSON format, which can be directly pushed to a database via AJAX.

Conclusions

We developed a web-based, platform-independent timing and counting application that allows researchers to easily record numerous tracking elements on a single screen, facilitating robust data collection for direct observational research.

Usability testing and comparative benchmarking against similar software are important next steps.

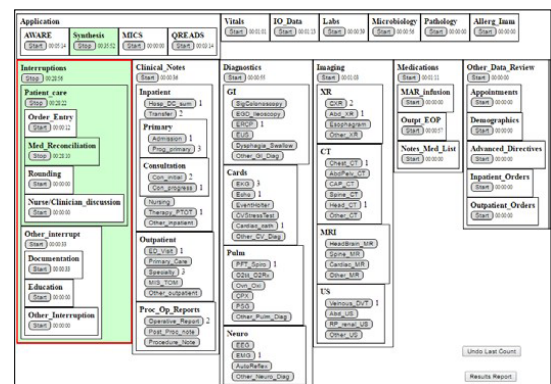


Figure 1 – DOctimer Application

Acknowledgements

No specific funding or conflicts of interest to declare.

References

- [1] M.A. Ballermann, N.T. Shaw, D.C. Mayes, R.T. Gibney, J.I. Westbrook, Validation of the Work Observation Method By Activity Timing (WOMBAT) method of conducting time-motion observations in critical care settings: an observational study, *BMC Med Inf Decis Mak* **32** (2011).
- [2] NoBox Inc, TimerData Plus, http://noboxinc.com/timer_data/index.html, Accessed December 13, 2016.
- [3] Macropod Software. ODLg. <http://www.macropodsoftware.com>, Accessed December 13, 2016.

Address for correspondence

Matthew Nolan: Nolan.Matthew@mayo.edu

Health Digital Divide in the Era of Precision Healthcare: Taking Yunnan Province as an Example

Yan Peng, Li Yan, Liuwang Kuanghua, Guozheng Tang

Institute of Medical Information of Yunnan Province, China

Abstract

With healthcare coming into the era of Precision Healthcare, it will facilitate the development of health informatics from information processing, integration, analysis, application and other uses. If we ignore the e-health development of developing areas, it will bring a health digital divide. This study takes Yunnan province as an example, and analyzes the problems of e-health development in a developing area, hoping to get more attention from researchers and regulators.

Keywords:

Digital divide, information services, delivery of health care

Introduction

Yunnan province is located in the southwest of China. The mountain area is about 94% of the total area. Due to economic, geographical conditions and other factors, information infrastructure, information technology application, information service capacity in Yunnan has lagged behind developed areas. In 2015, the Informatization Development Index of Yunnan was just 59.09, lower than the national average level (72.45), ranked 27th in 31 provinces [1]. This study analyzed problems of e-health development in Yunnan in the era of precision healthcare.

Methods

We investigated health administrators, health workers, patients by face-to-face interview in Zhenkang, Shangri-La, Yiliang, Wuding and Mengla, during October and November 2015. The interview outlines were determined by expert consultation method. Data from the interviews as well as information from literature review were aggregated and analyzed for themes.

Results

Interviews and literature review showed problems detailed below. First, some institutions, especially village clinics, cannot bear the cost of internet and computers, so that these institutions still do not access to internet. The health records created by village clinics cannot be accessed online. Second, data standards of information systems in the hospitals are not compatible, so it is very difficult to integrate and exchange of electronic medical records between hospitals. Third, health information systems in township hospitals are still at primary level, just for charge, financial management or pharmacy management. The electronic medical records are rarely used. The compatibility of health data is very poor. Fourth, because

of low level of education, aging, minority languages and other reasons, it is difficult for patients or health consumers to find accurate health information by internet in these areas.

Conclusions

There is a big gap between the condition of e-health development in Yunnan and the basic conditions of precision healthcare. In order to reduce the new health digital divide, policy makers and researchers must pay more attention to developing areas, information infrastructure of health institutions, health information literacy of health care workers and consumer needs in order to improve as soon as possible.

Acknowledgments

This study was supported by Rule of Law Division, People's Republic of China Committee on Health and Family Planning.

References

- [1] [Internet]. [cited 2017 June 2]. Available from: <http://www.miit.gov.cn/n1146290/n1146402/n1146445/c4838381/part/4838390.pdf>

Address for correspondence

Yan Peng, Institute of Medical Information of Yunnan Province, Building 5, Baoli Liuhe, West People Road, Kunming, Yunnan, China. Email: imiynpy@126.com

Measuring User Satisfaction with Clinical Information Systems: What Really Matters?

Jonathan L. Schaffer^a, Peter Haddad^{b,c}, Nilmini Wickramasinghe^{b,c}

^a Cleveland Clinic, Ohio, USA

^b Faculty of Health, Deakin University, Melbourne, Victoria, Australia

^c Health Informatics Management Unit, Epworth HealthCare, Richmond, Victoria, Australia

Abstract

The healthcare industry is increasingly investing in information systems / Information technology to enhance patient outcomes and organizational performance. This study qualitatively investigates the relationship between the overall satisfaction and five key aspects of clinical information systems. The results show that intuitive, easy-to-use, and collaboration enabling systems are more likely to satisfy users. The level of technical support and training also play key roles in determining user satisfaction in the clinical domain.

Keywords:

Information Systems, Personal Satisfaction, Health Care Behavior

Introduction

While examining IS/IT user satisfaction in healthcare has a lengthy history, measuring user satisfaction with clinical information systems lags behind. This is a major void in today's healthcare environment given significant investments in CIS (clinical information systems).

Aim

To examine user satisfaction of four clinical information systems focusing on five key functionalities; efficiency of use, intuitiveness of graphical user interfaces (GUI), communications, collaboration, and information exchange, and interoperability and compatibility issues.

Methods

An online survey was designed, validated and then conducted to collect data on clinical IT user satisfaction at a large tertiary, not-for-profit, private healthcare group in Australia. The objective of this survey was to develop a valid measurement tool of clinical IT user satisfaction that can be applied for all CIS given that currently no such measurement tool exists. Descriptive and predictive statistical techniques were employed to analyze the collected data.

Results

A total of 107 respondents out of 250 answered the questionnaire. Responses ranged from satisfied to not very

satisfied across the respective systems as well as frequency of use from multiple times a day to one or two times per month and varying technical proficiency. In addition, reluctance to contact IT help desk was a dominant response while significant differences between genders were not apparent but differences between level of proficiency and use were.

Discussion

Practically, the results of this study help executive decision makers including the CIO and CEO at hospitals to better understand the actual needs of clinical information systems' users to better utilize CIS as a contemporary asset as well as have more informed and critical discussion with prospective IT vendors. This is crucial with the increased investments in CIS. Today, healthcare is ranked fourth in investing in IS/IT after retail, banking and securities, and education. The study also shows that CIS users are likely to be satisfied if the systems are intuitive, easy to use, and enable better access to medical information in a timely manner. Moreover, results show that decision makers will need to pay attention to training and technical support channels. The amount and quality of training are key aspects of user satisfaction as the results show. From a theoretical stand point the developed assessment tool is one of the first to provide a comprehensive assessment of CIS user satisfaction.

Conclusions

This study set out to evaluate the overall user satisfaction with clinical information systems at an Australian tertiary, not for profit, private healthcare group. Different constructs were considered to evaluate the user satisfaction. The results show that intuitive, easy-to-use, and collaboration enabling systems are more likely to satisfy their users. The level of technical support and training seem to play key roles in determining user satisfaction in the clinical domain. Future research directions include fine tuning the survey to quantitatively determine user satisfaction based on its constructs in this study.

Address for correspondence

Corresponding author Prof. Nilmini Wickramasinghe, Epworth HealthCare and Deakin University.

Email: nilmini.work@gmail.com

Prerequisites for International Exchanges of Health Information for Record Research: Comparison of Australian, Austrian, Finnish, Swiss, and US Policies

Hanna Suominen^{a, b}, Henning Müller^c, Lucila Ohno-Machado^d,
Sanna Salanterä^{b, e}, Günter Schreier^f, Leif Hanlen^a

^a The Australian National University, Data61, & University of Canberra, Canberra, Australia, ^b University of Turku, Finland,

^c HES-SO & University of Geneva, Sierre/Geneva, Switzerland and Harvard Medical School, Boston, USA,

^d University of California, San Diego, USA, ^e Turku University Hospital, Turku, Finland,

^f AIT Austrian Institute of Technology, Graz, Austria

Abstract

The policies that address health information exchanges for research purposes in Australia, Austria, Finland, Switzerland, and the USA apply accountability and/or adequacy to protect privacy. Specific requirements complicate the exchanges: inform data subjects of data use purposes; assure that the subjects are no longer identifiable; destroy the data in the end; and not to use cloud computing without specific permission.

Keywords:

Electronic Health Records, Health Policy, Privacy

Introduction

A major risk in electronic health records is the possibility of compromising data subjects' privacy, and this is particularly evident in analyzing text (i.e., inability to be fully convinced that all privacy-sensitive information has been removed) or big data (i.e., unforeseen possibilities to infer personal data after record linkages from multiple de-identified sources). Using EHRs for research purposes requires compliance with legislation, and governance.

Methods

We specified the legal frameworks, process of gaining access to EHRs, and restrictions for data exchanges across projects in five countries: Australian Commonwealth, its NSW, Austria, Finland, Switzerland (Valais), USA, and its CA. We used a published method [1] and extended its analysis from Australia and Finland to the EU more widely (Austria), non-EU Europe (Switzerland), and N. America.

Results

Requirements for data access and protection vary (Table 1, [2]). The frameworks apply accountability of the original data creator for regulatory compliance (e.g., Australia and USA) and/or the subsequent information receiver having to protect privacy adequately (e.g., Australia and EU) (Table 2, [2]). ICT can audit compliance with all frameworks [3]. The process of gaining access to EHRs for research has five steps (Table 3, [2]): 1) Preparations include: developing a research

plan, research group, and an ethics protocol. 2) The proper approvals and permissions are furnished. 3) Data are collected and de-identified and an informed consent is obtained from each subject. 4) Research, where the exchanged data are used only for these purposes, takes place. Exchanges of the original or secondary data across borders or projects are permitted if they have been addressed in Steps 1-3; the use of cloud services, which may store data in another country or legislation, without specific permission is not allowed. 5) All data are deleted or returned to their original creator at the end.

Conclusion

Capabilities to exchange health information are critical to accelerate discovery and its diffusion to practice. However, the same ethical and legal policies that protect privacy hinder these exchanges. Both legislation and technologies are available for overcoming these barriers [3].

Acknowledgements

Profs Müller and Schreier, the EU's 7th framework program (Khresmoi and 261743); Prof Ohno-Machado by the US National Institutes of Health (U54HL108460 & UL1RR031980) and the US Agency for Healthcare Research and Quality (R01HS019913); and Prof Salanterä by the Academy of Finland (140323) and Tekes, the Finnish Funding Agency for Technology and Innovation (2227/31/2010). Adj/Prof Suominen received a travel grant by iDASH.

References

- [1] H. Suominen, Towards an international electronic repository and virtual laboratory of open data and open-source software for telehealth research: comparison of international, Australian, and Finnish privacy policies. *Stud Health Technol Inform*, **182** (2012), 153-160.
- [2] Available at <https://drive.google.com/file/d/0B58qEFIEoLAKNXhjOEJyVTNFZk0/view>
- [3] H. Suominen, H. Müller, L. Ohno-Machado, S. Salanterä, G. Schreier and L. Hanlen, Prerequisites for international exchanges of health information: Comparison of Australian, Austrian, Finnish, Swiss, and US privacy policies. <http://arxiv.org/abs/1612.04902> (2016).

Address for correspondence:

hanna.suominen@anu.edu.au

Key Factors for the Successful Adoption of IS/IT in Healthcare: A Fit-Viability Perspective

Nilmini Wickramasinghe^{a,b}, Peter Haddad^{a,b}, Imran Muhammad^{a,b}

^a Faculty of Health, Deakin University, Melbourne, Victoria, Australia

^b Health Informatics Management Unit, Epworth HealthCare, Richmond, Victoria, Australia

Abstract

The adoption of clinical information systems by healthcare organizations has become a strategic necessity today. For these situations to be successful, their fit becomes an important consideration. This research-in-progress study assesses the fit of three different clinical information systems in an Australian private not-for-profit healthcare group, and the readiness of this organization to adopt these systems. In so doing, it identifies key factors for the successful adoption of IS/IT in healthcare contexts. The study adopts a mixed methods approach using semi-structured interviews and an online survey.

Keywords:

Hospital Information Systems, Point-of-Care Systems, Telemedicine

Introduction

Adopting clinical information systems is becoming a strategic necessity for various levels of healthcare providers. The ultimate objective of most of these investments is to enhance patient outcomes with less cost. Many of these investments though, have not achieved the expected outcomes due to numerous factors as noted in the literature. These factors are mainly related to the readiness of healthcare organizations to adopt clinical information systems, and the fit of these systems into the organizational characteristics and objectives. This research-in-progress study assesses the fit of three different clinical information systems in an Australian private not-for-profit healthcare group, and the readiness of this organization to adopt these systems.

In order to assess the fit of these systems and readiness (viability) of the selected healthcare organization, this study uses the FIT-Viability Framework. The first system to be considered is an American developed computerized practitioner order entry (CPOE) system, the second is a European developed point-of-care system, and the third is an Australian made clinical auditing system. All of these systems are implemented by the selected case which operates in the Australian healthcare context.

Aim

The aim of this research is to build a framework that captures the factors that affect various clinical IS/IT solutions in the Australian healthcare context. In so doing, the study answers the following research questions:

1. What are the factors that affect the readiness of healthcare organizations to adopt sophisticated clinical IS/IT platforms?

2. How the domestication of internationally developed clinical IS/IT affect the fit of these systems in the Australian context?

Methods

This research is exploratory in nature, as it is planned to be a 'broad-ranging, purposive, systematic, and prearranged undertaking designed to maximize the discovery of generalizations leading to description and understanding of the area of research'. It includes both qualitative and quantitative methods and thus it is a mixed methods study. For the qualitative component, the study subscribes to the directives of Yin who notes that such an approach enables conducting in-depth studies and research about a broad range of topics and at the same time provides greater latitude in selecting topics of interest.

Further, the study includes 28 semi-structured interviews with key informants in a large not-for-profit tertiary healthcare group in Australia. This hospital represents an exemplary single case study with multiple units of analysis. The interviewees are classified into four categories based on their occupations and organizational tasks, namely management and executives, IT, clinicians, and clinical IT. The last group comprises clinical personnel with sound IS/IT knowledge and experience. Further, the study employs an online survey to target a wider audience at the selected case. Both the interviews and survey are designed to capture data about the fit of clinical IS/IT solutions used by the healthcare group, and the viability of the selected case to adopt these and other IS/IT solutions.

Conclusion

Currently, we have developed the conceptual model. Now we need to test it to unveil the factors that affect the fit of clinical IS/IT into an exemplary Australian healthcare provider and the viability of this hospital to adopt clinical IS/IT systems. This is particularly crucial given the current increasing trend to use IS/IT to deliver superior healthcare operations. We then plan to test the model in other healthcare systems including Canada, China, Germany and US. Once all stages of testing are complete, the robustness of the proffered model would be established and hence its benefits to improving the readiness of healthcare organizations and the fit of IS/IT systems globally will be the primary contribution of this study to theory and practice.

Address for correspondence

Corresponding author Professor Nilmini Wickramasinghe
Epworth HealthCare and Deakin University.
Email: Nilmini.work@gmail.com

Mobile Learning in Nursing Undergraduates in China: Current Status, Attitudes and Barriers

Qian Xiao^a, Qiannan Zhang^a, Lanlan Wang^a, Yanling Wang^a, Liu Sun^a, Ying Wu^a

^a School of Nursing, Capital Medical University, Beijing, China

Abstract

To explore the current status, attitudes and barriers of nursing undergraduates toward mobile learning, 157 nursing students were investigated. more than half of them used mobile learning frequently in past half year. The mean score of students' intention towards mobile learning was 10.5 (ranged from 6 to 15), and it related to students' gender, expected effect, ease of operation, influence of other students, self-learning management and perceived interest. Some barriers affected students' mobile learning. Therefore, students had positive attitude and perception toward mobile learning, then we should create enough conditions to promote students' mobile learning.

Keywords:

Attitudes, Barriers, Mobile learning, Nursing undergraduates.

Introduction

With the popularity of smart phones, mobile technology had been integrated into nursing education^[1]. Mobile learning, as a new form of learning, was effective, flexible, interesting and interactive, especially it can help students to learn anytime and anywhere^[2]. Hence, this study aimed to explore the current status, attitudes and barriers of nursing undergraduates toward mobile learning.

Methods

Students who had more than 1 year study experience in university were eligible to participate in this investigation. The questionnaire mainly includes 3 aspects: 1) Students' basic information, including age, gender, experiences of using smart phones and mobile learning. 2) Students' attitude towards mobile learning: including 6 domain and 18 items with 5-point Likert score (from 1 to 5), designed on the base of Aizhen Xie's UTAUT model of mobile learning^[3]. The 6 domain focused on the intention to mobile learning and its related 5 possible factors: expected effect, ease of operation, influence of students around, self-learning management and perceived interest. 3) Barriers to mobile learning.

Results

157 nursing undergraduates participated in this study which included 18 male and 139 female students. All of them were in the age group of 18-22 years, and 77.8% of them had more than 2 years experiences of using smart phones. 53.7% students used mobile learning frequently in past half year.

The score of students' intention towards mobile learning ranged from 6 to 15, with an average of 10.5 (SD, 2.3). The intention to mobile learning was related to students' gender, expected effect, ease of operation, influence of students

around, self-learning management and perceived interest (Figure1). The barriers to mobile learning were mainly reflected in the following three aspects: learning resources, features of smart phone and external support conditions (Figure1).

Discussion

This study showed that nursing students had the basis and positive intention for mobile learning, but decisive factors and barriers would affect their use of mobile learning. And it was clear what should pay attention to attract users for continuous adaptation and absorb more new users in developing mobile learning system and the mobile learning resources. It also contributes to the decision-making of teachers, and the formulation of policy on the promotion of innovative mobile learning.

Conclusion

Students had positive attitude and perception toward mobile learning, enough conditions to promote students' mobile learning should be provided.

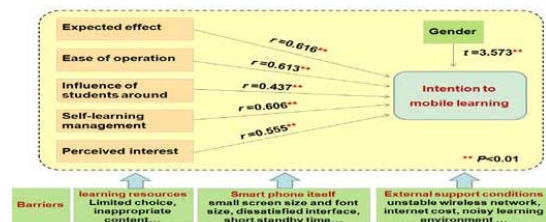


Figure 1 – Nursing undergraduates' intention to mobile learning and its related factors and barriers

References

- [1] B.J. Mackay, Anderson and T. Harding, Mobile technology in clinical teaching, *Nurse Educ Pract* 22(11) (2017), 1-6.
- [2] R.N. Patil, B.D. Almale, M. Patil and et al, Attitudes and Perceptions of Medical Undergraduates Towards Mobile Learning (M-learning), *J Clin Diagn Res* 10(10), (2016), 6-10.
- [3] X. Aizhen, Research on influence factors of university students' willingness of mobile phone learning base on UTAUT [D]. Zhejiang Normal University, 2012 (Chinese).

Address for correspondence

Corresponding author: Ying Wu, professor, PhD.

School of Nursing, Capital Medical University, Beijing, China.

Email: helenywu@vip.163.com. Phone: 86-10-83911766

Secure Electronical Communications and Data Transfers in a Clinical Environment

T. Xu^a, V.S. Thiemann^a, R. Röhrig^a, D. Brammen^b and R.W. Majeed^a

^aDepartment of Medical Informatics, Carl von Ossietzky University, Oldenburg, Germany

^bDepartment of Traumatology, Otto von Guericke University, Magdeburg, Germany

Abstract

As part of the German Project AKTIN, data security arises as an important issue. The context of this issue was detailed and the requirements were determined, with special focus on the hospital's point of view. The solution is illustrated in this poster. Further use cases in other medical or study context are also discussed.

Keywords:

Computer Security, Electronic Mail

Introduction

In modern day medical context, there are increasing needs to transmit data also outside of complex study software. Ransomware brought new awareness to IT-security in clinical context and a healthy scepticism towards emails. During the early stages in the German Emergency Department Data Registry (AKTIN)[1], anonymised data is exported for evaluation of data quality. These data exports need to be transferred from the hospitals to our data analyst. Discussions with IT staff at the hospitals have shown the following environment at the hospitals: The common operating system is Windows. New software installations are difficult and should be avoided. Software must be usable with only little IT knowledge. Computers may not have direct internet access. Sceptical users want to verify that only the intended data is transmitted. This Poster presents our solution for secure data transfers from a clinical environment.

Methods

There are several possible approaches from Computer Science to ensure data safety and authenticity during communication: GPG encrypted mails, diverse asynchronous or certificate based methods, and secured channel transport. In this instance, a hybrid encryption scheme is chosen: Symmetric encryption is used for the potentially large data payload, while the symmetric key itself is encrypted via asymmetric algorithms. The software solution consists of Windows-Batch-Scripts and bundled dependencies which can be run without installation.

Results

Since the file for transmission (called F) can be of variable size, a symmetrical encryption with a reasonable large, randomly generated key K_S is used, generating F_{enc} . The recipient's public key K_{Rp} is either stored in a local repository or retrieved from a trusted third server (with certificate) and used to encrypt K_S to K_{S-enc} . Both encrypted files are then

compressed and sent, possibly with the sender's signature, to the recipient. The recipient can then verify the sender by his public key and decrypt the K_{S-enc} with his own private key K_R . The encrypted file can be then decrypted with K_S .

With the hybrid encryption scheme, payloads of arbitrary length can be encrypted – independent of the asymmetric key block size. Transparency is provided by supplying not only the encrypted package, but also the original file and a software package for the physician to encrypt the file manually after checking the file to his satisfaction. All intermediate files are deleted at the end of encryption. The source code is available at <https://gitlab.uni-oldenburg.de/AKTIN/secure-transmit>

Discussion

The solution presented here will be used as an additional layer in the secured communication channels for the German Emergency Department Data Registry. There are other, already implemented methods in IT security fields. But these methods require additional software and management or are not fit for physician use, since they also require user training. Our solution provides data safety and prevents unauthorized data access. It also ensures data origin by signature. It can be extended or adapted for other projects and is easy to integrate into existing software. The only requirement is a trusted server. The possible senders and receivers need to create public and private key pairs and either register their public keys with the trusted server or exchange them with each other.

Conclusions

We presented a file encryption system by means of a small software package usable for most hospital IT environments, which improves data security compared to the commonly used unprotected channels.

Acknowledgements

Funding by BMBF No. 01KX1319A, 01KX1319B

References

- [1] J. Ahlbrandt, D. Brammen, R.W. Majeed, R. Lefering, S.C. Semler, S. Thun, F. Walcher, and R. Rohrig, Balancing the need for big data and patient data privacy – an IT infrastructure for a decentralized emergency care research database, *Studies in health technology and informatics* 2014, 205:750–4.

Address for correspondence

Tingyan Xu, tingyan.xu@uol.de

This page intentionally left blank

IV. Knowledge Management

This page intentionally left blank

Big Data in the ICU: experience in the Hospital Italiano de Buenos Aires

José Astudillo^a, Javier Zelechower^a, Francisco Traversaro^a, Francisco Redelico^b, Daniel Luna^c, Fernan Quiros^c, Marcelo Risk^{b,c}, Eduardo San Román^a

^aService of Adult Intensive Care Unit, Hospital Italiano de Buenos Aires, Argentina

^bConsejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

^cDepartment of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

Intensive care represents the critical care setting of a hospital, where fundamental, precise, and fast decisions have to be made. These decisions will affect the outcome of the patients in a matter of few hours. The knowledge of the therapeutic interventions applied in this setting is evolving, thus the perspective of Big Data may provide a new paradigm in the ICU. The conformation of a multidisciplinary team is essential to develop Big Data in the ICU.

Keywords:

Intensive Care Unit, Critical Care

Introduction

Big Data is defined in many ways, especially given its origin in the economic and social sciences. The interest and application to the medical sciences is promising given the potential to generate a high volume and variety of data, and new knowledge in accordance with this abundance of information. In broad outline, and as other authors did, Big Data can be broken down into three fundamental aspects: Volume, Speed and Variety [1]. This approach is little known in the medical field and require the work of multidisciplinary teams, for their interpretation and application to clinical practice [2]. The objective of this poster is to propose a system of analysis of the vital signs and data of the Electronic Health Record (EHR) in critical patients.

Methods

The Hospital Italiano Buenos Aires has a mature EHR, developed entirely in our institution. All the interventions and tests of the patients are registered, regardless of the reason for consultation as well as the degree of complexity required in their care. We propose the formation of a multidisciplinary team composed of physicians, engineers and biologists, in order to develop this new approach of managing medical information.

This approach of using extensive medical information, which occurs during the hospitalization in the intensive care unit, the patients data is collected during their first 24 hours of hospitalization in the ICU. The sample of selected patients included immediate postoperative patients and shock patients with suspicion of septic etiology. It was collected from this population two severity scores Acute Physiology And Chronic Health Evaluation II, vital support requirements (vasopressors, mechanical ventilation, and sedation), and balance for the first 24 hours. It was also recorded, the duration of stay in both the

ICU and the hospital, and infectious inter occurrence during hospitalization.

The pilot study have a total of 20 patients. We delight a total of 120 patients for the next 6 months. After that, we will process all the gathered information in order to obtain patterns of behavior towards sepsis. We expect this patterns will help us predict other patients evolution during their internment.

Discussion

The basic characteristics of Big Data would allow to, in the context of the critical patient, produce more accurate diagnostic impressions, based on the signals analysis from the vital signs and the records in the electronic medical record. In turn, the diagnostic impressions will lead, through algorithms based to a known institutional population, on the most relevant diagnostic tests and procedures. In this way, it would be possible to optimize resources and adjust the best procedures for any given patient.

In conclusion, this approach, combined with the genetic content of the patients, will allow the practice of precision medicine in critical patients [4]. Finally, through the knowledge of the outcome of similar patients undergoing similar treatments, Big Data could help to choose the most appropriate interventions. This is, to the extent of our knowledge, the first experience of this type of a multidisciplinary team work in an Argentinian Hospital.

References

- [1] Gandomi, A., and Haider, M., Beyond the hype: Big data concepts, methods, and analytics, *Int. J. Inf. Manage.* **35** (2015) 137–144.
- [2] Murdoch, T. B., and Detsky, A. S., The Inevitable Application of Big Data to Health Care, *JAMA* **309** (2013) 1351.
- [3] Anthony Celi, L., Mark, R. G., Stone, D. J., and Montgomery, R. A., “Big Data” in the Intensive Care Unit. Closing the Data Loop, *Am. J. Respir. Crit. Care Med.* **187** (2013) 1157–1160.
- [4] Buchman, T. G., Billiar, T. R., Elster, E., Kirk, A. D., Rimawi, R. H., Vodovotz, Y., et al., Precision Medicine for Critical Illness and Injury, *Crit. Care Med.* **44** (2016) 1635–1638.

Address for correspondence

Marcelo Risk, PhD, ScD. Hospital Italiano de Buenos Aires, Buenos Aires, Argentina. E-mail: marcelo.risk@hospitalitaliano.org.ar

Training of Students for Critical Evaluation of Mobile Health Applications

Jadranka Bozikov^a, Mira Hercigonja Szekeres^b

^a Andrija Stampar School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia,

^b University of Applied Health Studies, Zagreb, Croatia

Abstract

Medical and nursing students need to be trained in critical evaluation of mobile health applications (m-health apps) in order to see their potential and to understand the need for their validation and certification. Students should develop positive attitudes but also critical views on such apps in order to be prepared for responsible and ethical use of m-health in their future professional practice.

Keywords:

Education, Medical; Mobile Applications; Telemedicine

Introduction

Health and social care systems are facing an upsurge in chronic diseases and multi-morbidity as a consequence of population ageing, at least when the European Union and broader European Region is considered. There is a growing need for supportive tools and practices that will enable elderly people to be active and independent. Mobile applications for health (m-health apps) have been recognized for their great potential to improve the health and wellbeing of patients by empowering them to take responsibility for their own health [1,2]. Likewise, m-health apps hold a promise for innovative health promotion interventions targeted at younger generations. Therefore, health professionals need to take into account the advantages of m-health apps, but also to be aware of their limitations and even harmfulness. Students need to be trained in critical evaluation of m-health apps in order to be prepared to utilize the potential of m-health in their future professional practice.

Methods

Fifth-year medical students and second year Master programme nursing students were asked to download m-health apps for smartphones and test them according to given guidelines. Each student has to review and report (in the form of an oral presentation accompanied with PowerPoint) on the selected application regarding the following points:

- target population: patients or healthy subjects
- aim and function: information and education only / health promotion / health care and disease control
- need of Internet connection: all the time or occasionally
- possibility of connecting measuring and tracking units (e.g. heart rate, Fitbit, belt, watch)
- communication with other persons: caretakers, healthcare personnel or social networks

- authors/publishers, availability, languages, price, etc.

Results

Students have proved their capability for testing m-health apps and understand the need for their evaluation, validation and certification. They successfully documented test results, presented them during seminars and have discussed the apps' advantages and drawbacks, as well as their potential, from the perspective of health care.

Conclusions

Medical and nursing students should be trained in critical evaluation of m-health applications in order to:

- develop positive attitudes towards their use, but also critically view and understand the necessity for their validation and certification
- be prepared for the responsible and ethical use of m-health apps in future professional practice

References

- [1] Blueprint Digital Transformation of Health and Care for the Ageing Society. v. 15 January 2017. Available at: http://ec.europa.eu/newsroom/document.cfm?doc_id=40787
- [2] Mobile Medical Applications. Guidance for Industry and Food and Drug Administration Staff. Document issued on February 9, 2015. Available at: <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM263366.pdf>

Address for correspondence

Jadranka Bozikov, e-mail: jbozikov@snz.hr

Developing the First Generally-Available openEHR Archetypes and Templates for Physiotherapy: An Example of Building Clinical Models and Modelling Capacity via Student-Led Academic–Industrial Collaboration

Jamila Chihab^a, Hildegard Franke^b, Ian McNicoll^{a,b}, Matthew W. Darlison^a

^aInstitute of Health Informatics, University College London (UCL), London, UK

^bfreshEHR Clinical Informatics Ltd, Kettering, Northamptonshire, UK

Abstract

We present the first public openEHR archetypes and templates for physiotherapy, and the context of multidisciplinary academic-industry partnership that has enabled their production by a team led by a clinically trained student on the UCL health informatics MSc programme.

Keywords:

Electronic Health Records, Physical Therapy Specialty, Computer Simulation

Introduction

Given the clear assertions in published literature that physiotherapy needs, and will benefit from, electronic health records, it was surprising that the international repository for clinical models for the openEHR record architecture (the *Clinical Knowledge Manager* or CKM at openehr.org/ckm) made no reference to physiotherapy, and that other published modelling work had not used formalisms that provided for easy re-use. We set out to explore what would be involved in changing this.

Methods

With advice from another domain expert, a guideline for the assessment of ankle sprain by a physiotherapist [2] was selected. The guideline was chosen to be both generic (i.e. not from a specific hospital) and informed by established principles of structured record keeping (the *International Classification of Functioning, Disability and Health*).

Close reading of the guideline text enabled identification, extraction and enumeration of all data items mentioned or implied, and their datatypes. This included the tracing of recording requirements included in the guideline by reference (e.g. to particular tests or scoring systems).

The resulting document informed a “pair modelling” (cf. *pair programming*) process in which the student and an industry expert modeller worked side by side to organise the guideline content and build a corresponding openEHR template for ankle sprain assessment backed by existing and, where necessary, new openEHR archetypes.

The template and archetypes were then published on the UK national CKM (clinicalmodels.org.uk) for review, thus also making them available internationally for scrutiny and possible adoption by others.

Results

The *Ankle Sprain - Assessment UCL.v0* template uses 41 archetypes, 25 of which were drawn from the international CKM. The CKM review was largely positive, raising some technical modelling questions and identifying some flaws in the original guideline.

Discussion

Imperfections notwithstanding, the real-world utility of the artefacts produced supports the hypothesis that clinical modelling is a field in which a student domain expert can – with appropriate support – develop sufficient expertise to make credible contributions to a public-facing knowledge resource.

Conclusions

Guideline authors should consider (and perhaps even specify) the consequences of their editorial decisions in relation to the record-keeping required in a multiprofessional, multi-agency service context. Taking a guideline, which already has some degree of clinical consensus, as a starting point significantly eases the process.

This kind of academic-industry partnership is rewarding for participants and aligns well with research-based education, for example as articulated in UCL’s Connected Curriculum [1].

Acknowledgements

We thank Dr Ibtissam Saab for expert advice on guideline selection and all who contributed time and expertise to the review process.

References

- [1] D. Fung, *A Connected Curriculum for Higher Education*, London: UCL Press, 2017.
- [2] M. Granado and D. Matlick, Clinical Review: Ankle sprain, in: *Rehabilitation Reference Center* Cinahl Information Systems, 2016.

Address for correspondence

Dr Matthew W. Darlison (m.darlison@ucl.ac.uk)
Institute of Health Informatics, University College London,
Gower Street, London WC1E 6BT, UK

Exploiting Temporal Constraints of Clinical Guidelines by Applying OpenEHR Archetypes

Lilian Mie Mukai Cintho^a, Diego Garcia^a, Bruno Henrique da Silva Santos^a,
Lucia Sacchi^b, Silvana Quaglini^b, Claudia Maria Cabral Moro^a

^aHealth Technology Graduate Program, Pontifícia Universidade Católica do Paraná, Brazil

^bDepartment of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

Abstract

Studies describing Computer-Interpretable Clinical Guidelines (CIG) with temporal constraints (TC) generally have not addressed issues related to their integration into Electronic Health Record (EHR) systems. This study aimed to represent TCs contained in clinical guidelines by applying archetypes and Guideline Definition Language (GDL) to incorporate decision support into EHRs. An example of each TC class in the clinical guideline for management of Atrial Fibrillation was represented using archetypes and GDL.

Keywords:

Decision Making, Computer-Assisted; Decision Support Systems, Clinical; Decision Support Techniques

Introduction

Researchers have addressed issues related to the representation of clinical guidelines with TC specification using formalisms and tools to develop clinical decision support systems (CDSS). Although it is difficult to integrate them into the EHR, it is necessary to map the information. The integration of CDSS into

EHR based in the archetypes architecture that is a standard to system interoperability and make possible to represent data and rules of CDSS, avoiding the mapping, was presented by [1], however not dealing with TC.

Accordingly, this work aims to represent the TCs contained in clinical guidelines with archetypes, and to use the GDL as a unifying language for representing both the EHR and CDSS.

Methods

An example from each type of TC class was identified: qualitative, quantitative and mixed, from the European Guideline for the Management of Patients with Atrial Fibrillation (AF). "For patients with AF of ≥ 48 h duration, or when the duration of AF is unknown, OAC therapy is recommended for ≥ 3 weeks prior to and for ≥ 4 weeks after cardioversion, regardless of the method (electrical or pharmacological)", is an example of a mixed (qualitative and quantitative) TC. The data and rules related to each TCs were represented using archetype and GDL. The execution of each rule was validated manually, with 20 real cases from patients randomly selected.

Figure 1 - Rules to represent TC in quantitative duration

Results

All the results of the rules execution are the same compared to the medical decision. Figure 1 shows part of the GDL rule created to represent a mixed TC.

Conclusion

The OpenEHR and GDL archetypes allowed to represent TCs contained in the CG.

Acknowledgements

CAPES (Brazilian Coordination for the Improvement of Higher Education Personnel) for the scholarship.

References

- [1] R. Chen, C. Valladares, I. Corbal, N. Anani, S. Koch, Early experiences from a guideline-based computerized clinical decision support for stroke prevention in atrial fibrillation, *Stud Health Technol Inform* 192 (2013), 244–247.

Correspondence Address:

miemukai@hotmail.commiemukai@hotmail.com

Towards the Consideration of Diagnostic Delay in Model-Based Clinical Decision Support

Jan Gaebel, Mario A. Cypko, Steffen Oeltze-Jafra

Innovation Center Computer Assisted Surgery, Faculty of Medicine, University Leipzig, Leipzig, Germany

Abstract

Diagnostic delay involves the peril of information becoming outdated. It is a challenging task to quantify the up-to-dateness of clinical information and the consequences of diagnostic delay with the goal of considering them in clinical decision support. We propose an approach to integrating the up-to-dateness of clinical information in a model-based therapy decision support system.

Keywords:

Decision Support Systems, Clinical; Models, Statistical; Cancer Staging

Introduction

The *diagnostic delay* refers to the duration of the diagnostic process until the final treatment decision is made, e.g. for suspected tumors. Gómez et al. found a significant correlation between diagnostic delay and advanced tumor stages [1]. We propose 1) a temporal rating of measured diagnostic delays and evaluation of the up-to-dateness of clinical information as well as 2) the integration of this rating in a model-based decision support system (DSS) for the staging of laryngeal cancer.

Methods

We exemplarily analyzed 39 patient cases with laryngeal cancer from 2013. The median delay for neck CT scans is 35 days, for the histopathological reports the delay is 6 days. Data beyond this temporal threshold should have less impact on the therapy decision. To answer the question of how outdated information should be processed and in what scenario this would be applied, we integrate a temporal rating of measured delays into a clinical DSS using Bayesian networks (BN). A BN has the potential to represent complex decisions by a model including all relevant patient characteristics and decision options. In the model, characteristics and decisions are represented by variables, each with a number of possible states, see Fig. 1. A decision model can be instantiated with an amount of certainty reflecting the real-world situation [2]. Variables' states can be instantiated, either with 100% certainty for one state (Fig. 1, left) or with probabilities distributed over multiple states (Fig. 1, right). They must always add up to 100%. If information is not given, this can also be specified, e.g. with an "unknown"-state. We employ the unknown state to reflect the uncertainty caused by diagnostic delay. For an outdated information, we set its probability to a value >0% depending on the temporal rating. For an initial analysis, we propose to set the uncertainty to 50% to halve the impact of the outdated information.

Results

Our approach connects the hospital information system with the therapy DSS for laryngeal cancer staging [3]. For testing purposes, we investigated a complex patient case: patient with 40 days old CT data and five days old pathological data. The DSS falsely computed a TNM staging of "T2 Nx Mo". The CT data, in which no infiltration of lymph nodes was visible, interfered with the more current pathological data. With the reduction of certainty of the outdated CT data, as illustrated in Fig. 1, right, the DSS correctly calculated the stage "T2 N2b M0". Setting "unknown" to 100% however gave no improvement. This shows that information cannot be discarded completely.

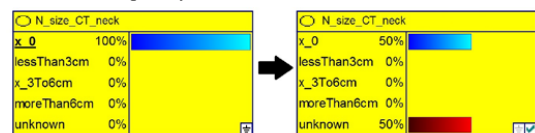


Figure 1 – Adaptation of certainty values for outdated data

Conclusion

We conclude that this approach would be easily applicable to other diseases or problems with the proper amount of medical knowledge and modeling efforts.

We will extend the calculation of the delay to all relevant diagnostic modalities to support the entire process and increase the quality of care. We will conduct a validation study on a larger patient data set.

References

- [1] I. Gómez, et al., Is diagnostic delay related to advanced-stage oral cancer? A meta-analysis, *Eur. J. Oral Sci.*, vol. 117, no. 5, pp. 541–546, Oct. 2009.
- [2] M.A. Cypko, M. Stoehr, A. Dietz, H.U. Lemke, A CDSS based on MEBNs with an example of therapy decision making for laryngeal cancer, presented AHNS, Seattle, WA, USA, 2016.
- [3] J. Gaebel, M. A. Cypko, and M. Stoehr, "Integrating Intelligent Agents in form of Arden Syntax for Computing Instances Based Fuzziness in Patient Specific Bayesian Networks," in Proceedings of the Int Joint Conf CARS, Heidelberg, Germany, 2016, vol. 11.

Address for correspondence

Jan Gaebel (jan.gaebel@medizin.uni-leipzig.de)
University Leipzig, Faculty of Medicine
Innovation Center Computer Assisted Surgery (ICCAS)
Semmelweisstr. 14, 04103 Leipzig, Germany

Considerations of Human Factors in the Design and Implementation of Clinical Decision Support Systems for Tumor Boards

Carmelo Gaudioso^a, Peter Elkin^b

^aDepartment of Biostatistics, Roswell Park Cancer Institute, Buffalo, New York, USA

^bDepartment of Biomedical Informatics, State University of New York at Buffalo, Buffalo, New York, USA

Abstract

Critical decisions are made at tumor boards (TB). Further research is needed to guide the design and implementation of clinical decision support systems (CDSS) for highly interactive, fast-paced, and complex TB decision making. We highlight two key human factors that should be considered in the design and Implementation of CDSS for multidisciplinary deliberations: cognitive burden and trust.

Keywords:

Decision Support Systems, Clinical; Decision Making; Cognition

Introduction

TB are considered a standard of care and the use of information systems is recommended to support evidence-based decision making [1]. Multiple human factors must be considered in the design of CDSS for complex and fast-paced multidisciplinary decision making. To accomplish this, requires a shift in the approach to the design of information systems [2]. Tumor board systems need to present complex information from across many disciplines to a clinician treating the patient. In order to assure that human factors are taken into consideration, the interface design is going through much iteration of usability studies. The dynamic interactions between system and users, and communicative actions are key considerations in the design of actable CDSS [3].

Methods

Following a qualitative research method, including review of archival data, observation, interviews, surveys, and researchers' reflection and discussion, through an iterative and incremental process [4], we designed an instance of a sub-class of CDSS for multidisciplinary decision making at breast cancer TB meetings. During testing, we observed and examined tumor board deliberations to characterize form, content, functional requirements, workflow, cognitive processes, and patterns of use of case-relevant evidence-based information with and without the aid of the CDSS. Semistructured interviews provided for more indepth and targetted analysis of our observations yielding rich data.

Results

Given the limited time to review each case, users required that only case relevant clinical information be displayed on a single screen, organized in a case (problem) relevant and logical format. However, Radiology and pathology reports were to be

displayed in their entirety rather than as a summary of findings. To this point, a medical oncologist said: "I want to read the report by myself. I don't trust somebody else's interpretation. These reports contain a set of critical information with details that are of different interest to different disciplines and tasks that have to be performed. Structure and synthesis are good for displaying discrete data (facts, i.e., clinical findings and laboratory results), but not for narratives like pathology and radiology reports."

Conclusion

Human factors, including cognitive load, trustworthiness and workflow inherent to dynamic and complex multidisciplinary decision making are critical considerations in the design of CDSS [5]. The design process must be one in which there is open communication and collaboration between the designer and the user and usability testing in real action bring to light critical form, content, and functional requirements.

Acknowledgements

The authors acknowledge the contributions of the tumor board members.

References

- [1] B.N. Doebbeling, A.F. Chou AF, and W.M. Tierney, Priorities and strategies for the implementation of integrated informatics and communications technology to improve evidence-based practice. *J Gen Intern Med* **21**(2006), S50–7.
- [2] A. Hevner, S. March, J. Park, and P.L. Rees, Design science in information systems research. *MIS Quarterly* **28**(1) (2004), 75–105.
- [3] P.J. Agerfalk and G. Goldkuhl, Business action and information modeling: The task of the new millennium. In Rossi, & Siau (Eds.), *Information modeling for the new millennium*, 2001.
- [4] R. Sorensen, A comparison of software development methodologies, *CrossTalk* **8**(1), 1995.
- [5] P.L. Elkin, Human Factors Engineering in HI: So What? Who Cares? and What's in It for You? *Health Inform Res* **18**(4) (2012), 237–241.

Address for correspondence

Carmelo Gaudioso, Roswell Park Cancer Institute, Elm & Carlton Streets; Buffalo, New York 14263.
Carmelo.Gaudioso@RoswellPark.org.

Usability Study of RSNA Radiology Reporting Template Library

Yi Hong^a, Ying Zhu^b, Selen Bozkurt^c, Jin Zhang^b, Charles E. Kahn, Jr^d

^a MedaAnalytics, Inc. Emeryville, California, USA

^b University of Wisconsin- Milwaukee, Milwaukee, Wisconsin, USA

^c Akdeniz University, Konyaaltı, Antalya, Turkey

^d University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

This study provides insights that could help to improve the Radiological Society of North America (RSNA) Reporting Template Digital Library, based on a usability evaluation. The results show that most users have been satisfied with the website. The general comments for the library are positive, although the participants suggested quite a few areas to improve. About 40% are returning visitors which means people often come back to the website.

Keywords:

Radiology Information Systems; Libraries, Digital; Program Evaluation.

Introduction

The Radiological Society of North America (RSNA) Radiology Reporting Template Library is a freely accessible digital library that incorporates metadata, standardized medical terminologies and health care industry standards to promote knowledge sharing of structured radiology reports [1]. Since the library was launched in December, 2009, it has received more than 4.2 million page views with its fast-growing collections of 252 reporting templates. The purpose of this study is to provide insights to improve the reporting template library website based on a usability evaluation, including an online survey, focus groups, web analytics, and transaction log analysis.

Methods

To get first-hand feedback from users, an ongoing online survey was posted on the library website. The web-based Qualtrics survey software was used to create the ongoing survey at https://milwaukee.qualtrics.com/jfe/form/SV_eR6Yawd3H_xiT_NFH. Two focus groups were formed to further gather users' feedback on the template library. One group consisted of three medical informaticians, while the other group included three radiologists. A protocol was designed for the focus group interviews. The focus group sessions were conducted while the results were summarized and interpreted in a report. A Google Analytics account was set up to track the usage of the website, as well as monitor the number of visitors and their geographical location. The scripts that enable Google Analytics to collect data were added to the website to track users' interactions with the website for a specific time. Approximately one year's data from the web analytics and transaction logs were analyzed using the Web Log Explorer to further identify demographic profile of visitors, determine user behavior as related to the library website, find out what search terms were frequently used to find templates, and determine what templates are most popular.

Results

A total of 163 respondents participated in the online survey from February 2016 to April 2017; 135 completed all the questionnaires. The respondents were from Asia (35.58%), North America (31.90%), Europe (16.56%), and Africa (9.82%). Most of the respondents were male (63%). The survey results suggested that most users had been satisfied with the reporting template library website. About 80% of survey respondents rated the overall level of satisfaction for the library website as excellent/good, while more than 70% respondents rated each of the following specific features as excellent/good: Navigability 76%, Timeliness 75%, Ease of search 71%, Availability of information 70%, Organization 80%, Visual appeal 76%, and Uniqueness 76%.

The focus group interview results showed that the general feedback for the library was favorable, although the participants suggested quite a few areas for improvement. The participants thought that the goal of digital library was remarkable, while the quantity of templates was adequate. The medical informatician group had a higher overall level of satisfaction with the website, while the radiologist group seemed much pickier with the content of the templates. Both groups thought the online library tours were not simple enough to show how to use the template library and suggested that a video tutorial should be used. The results of web analytics showed that most visitors to the library website resided in the United States. In terms of visitor profiles, a larger proportion were new visitors (61.4%) vs. returning visitors (38.6%) and 12,246 of 113,618 users (10.8%) were active users. As of April 6, 2017, the library website received 4,277,407 page views. The top five specialties that got most page views were Musculoskeletal Radiology, Ultrasound, Computed Tomography, Chest Radiology, and Diagnostic Radiology. There were total 684 search terms that occurred from transaction log analysis between 01/17/2016 to 04/07/2017.

Conclusion

This usability study provided insight into the evaluation of the reporting template library website, both methodologically and practically. It is the first study to perform a usability assessment for the radiology reporting template library. Our methodology could serve as an example for measurement development of other services in radiology reporting.

Reference

- [1] Radiology Reporting Template Library, Radiological Society of North America, available at <http://www.radreport.org>

Address for correspondence

Yi Hong, MedaAnalytics, Inc, 5858 Horton Street, Emeryville, CA 94608, USA, yiphong@yahoo.com

Evaluation of SNOMED CT as a Reference Terminology for Standardized Data Queries in the Arden Syntax

Robert A Jenders, MD, MS, FACP, FACMI

Charles Drew University & University of California, Los Angeles, California, USA

Abstract

Context: Arden Syntax is a standard that encodes knowledge as Medical Logic Modules (MLMs) but lacks a standard query data model and terminology. *Objective:* Assess to what extent SNOMED CT can represent MLM query data elements. *Method:* 340 MLMs containing 3268 queries were examined. *Result:* SNOMED can be used to represent all data elements except for clinical decision support system terms such as “alert.” *Conclusion:* SNOMED CT adequately represents data queried using the Arden Syntax.

Keywords:

Clinical Decision Support Systems

Introduction

Arden Syntax is a formalism supervised by Health Level Seven International (HL7) for representation of procedural medical knowledge to facilitate sharing units of knowledge known as MLMs that are executed in clinical decision support systems (CDSS). Some site-specific changes must occur in order for a knowledge base to be shared. The key to minimizing changes is the standardization of database linkages, which in turn requires at least a standard data model, terminology and query syntax. This is known as the “curly braces problem” of Arden because of the syntactic construct used to enclose these site-specific references [1]. Prior work [1] has established the utility of standards such as FHIR to serve as the standard data model for the Arden Syntax. SNOMED CT is a reference terminology used to encode health care data that, as such, can be used to standardize references to data used in knowledge interventions that can be represented as MLMs. The present work was undertaken to assess the utility of SNOMED CT as a reference terminology for queries in the Arden Syntax that, in conjunction with a standard data model, will help solve the curly braces problem.

Methods

A previously assembled convenience sample of MLMs was examined. The data query statements were extracted from these MLMs, and the data elements therein were identified. Each was assessed to ascertain whether it could be encoded using the September 2016 US Edition of SNOMED CT.

Results

A total of 340 MLMs were pooled from 5 source CDS systems, including 24 from 2 vendor knowledge bases and 316

from 3 academic medical centers. MLMs concerned with lab tests were the most common (138/340 = 41%), followed by clinical assessment (75=22%) and medication (45=13%). The remainder addressed administrative and miscellaneous topics. Each MLM contained at least one READ statement with a data query. A total of 3268 queries were identified, and the data elements therein were compared to concepts in SNOMED CT to assess if the former could be encoded by the latter. All the data elements in these queries could be encoded in SNOMED CT except for data elements referring specifically to aspects of CDSS (203 or 6.2% of all queries) such as alerts and intermediate blackboard-type results of preprocessing of data. Of note, the primary time of query data elements in Arden is an implicit attribute of every query variable and is not explicitly represented, but this can be encoded in SNOMED CT.

Conclusions

SNOMED CT is adequate to encode the data elements found in a large set of query statements in a corpus of Arden Syntax MLMs. The addition of concepts to SNOMED CT to represent CDSS constructs would improve its utility for this purpose. Consideration should be given for use of SNOMED CT to encode query data elements in a standard way in the Arden Syntax in order to facilitate knowledge sharing.

Acknowledgements

This work was supported by NIMHD grants U54MD007598 and S21MD000103 and NCATS grant UL1TR001881 from the National Institutes of Health (USA).

References

- [1] Jenders RA, Adlassnig KP, Fehre K, Haug P. Evolution of the Arden Syntax: key technical issues from the standards development organization perspective. *Artif Intell Med* 2016;pii: S0933-3657(16)30239-1.

Address for correspondence

Robert A. Jenders, MD, MS, FACP, FACMI
Center for Biomedical Informatics, Charles Drew University
Department of Medicine, University of California, Los Angeles
1731 E 120th St, Los Angeles, CA 90059 USA
jenders@ucla.edu

Building Interoperable FHIR-based Vocabulary Mapping Services: A Case Study of OHDSI Vocabularies and Mappings

Guoqian Jiang^a, Richard Kiefer^b, Eric Prud'hommeaux^b, Harold R. Solbrig^b

^aDepartment of Health Sciences Research, Mayo Clinic, Rochester, MN, USA

^bW3C/MIT, Boston, MA, USA

Abstract

The OHDSI Common Data Model (CDM) is a deep information model, in which its vocabulary component plays a critical role in enabling consistent coding and query of clinical data. The objective of the study is to create methods and tools to expose the OHDSI vocabularies and mappings as the vocabulary mapping services using two HL7 FHIR core terminology resources *ConceptMap* and *ValueSet*. We discuss the benefits and challenges in building the FHIR-based terminology services.

Keywords:

Reference Standards; Observational Study; Vocabulary, Controlled.

Introduction

The Observational Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM) is a deep information model [1] that specifies how to encode and store clinical data at a fine-grained level, ensuring that the same query can be applied consistently to databases around world.

HL7 Fast Healthcare Interoperability Resources (FHIR) [2] is emerging as a next generation standards framework for facilitating health care and electronic health records (EHRs) data exchange. The objective of the study is create methods and tools to expose the OHDSI vocabularies and mappings using the FHIR-based terminologies services.

Methods

We examined both the OHDSI Vocabulary CDM version 5.0.1 [3] and the STU3 Ballot version of the FHIR core terminology resources [2]. We created mappings between the Vocabulary CDM and two FHIR terminology resources with high maturity level – *ValueSet* and *ConceptMap*. We used FHIR extension mechanism to capture those fields in CDM (e.g., domain_id, vocabulary_id) that do not have corresponding mappings in FHIR. We installed an OHDSI virtual machine (VM) based on the CDM 5.0.1 version, which contains the full OHDSI Technology Stack and is loaded with the standardized vocabularies. We then created a Java-based transformation tool that invokes the HAPI-FHIR API to transform the OHDSI vocabularies and mappings as the instances of the FHIR *ValueSet* and *ConceptMap*, and loaded the instances into a local FHIR server.

Results

The OHDSI VM is loaded with 3,316,702 unique concept ids, in which 1,052,060 concepts are marked as “standard concepts” which have mappings asserted. These concepts are

classified by 32 domain ids, 61 vocabulary ids and 210 concept class ids. We have successfully created tools to generate the *ValueSet* instances that capture the metadata of concepts and the *ConceptMap* instances that capture the metadata of concept mappings (an example in Figure 1). We loaded 1000 *ConceptMap* instances and 2000 corresponding *ValueSet* instances in a FHIR server and examined the search capability enabled by FHIR.

```
<group>
  <source value="http://ohdsi.org/cdm/ValueSet/313217-standard-concept"/>
    <sourceVersion value="ohdsi-cdm-v5"/>
    <target value="http://ohdsi.org/cdm/ValueSet/313217-mapped-from"/>
    <targetVersion value="ohdsi-cdm-v5"/>
    <element>
      <extension url="http://ohdsi.org/cdm/concept/conceptId">
        <valueString value="313217"/>
      </extension>
      <extension url="http://ohdsi.org/cdm/concept/vocabularyId">
        <valueString value="SNOMED"/>
      </extension>
      <code value="49436004"/>
      <target>
        <extension url="http://ohdsi.org/cdm/concept/conceptId">
          <valueString value="44821957"/>
        </extension>
        <extension url="http://ohdsi.org/cdm/concept/vocabularyId">
          <valueString value="ICD9CM"/>
        </extension>
        <code value="427.31"/>
        <equivalence value="equivalent"/>
      </target>
    </element>
  </group>
```

Figure 1. A segment of a *ConceptMap* instance for a mapping between two codes from SNOMED CT and ICD9-CM.

Discussion

The OHDSI Vocabulary CDM and its implementation have been successfully used in supporting clinical observational data integration and systematic data characterization. It has provided a collection of valuable terminology service requirements that can be generalizable to address the similar needs of broader scientific communities. The FHIR terminology resources and tooling provide a standard mechanism to enable interoperable vocabulary mapping services for the OHDSI vocabularies and mappings.

Acknowledgements

This study is supported in part by NIH grants U01 HG009450, U01 CA180940, and R01 GM105688.

References

- [1] Hripcsak G, et al. Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci U S A*. 2016 Jul 5;113 (27):7329-36.
- [2] HL7 FHIR. <http://hl7.org/fhir/2016Sep/index.html>. Last visited at November 24, 2016.
- [3] OMOP CDM V5.0.1: <http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm:single-page>. Last visited at November 24, 2016.

Address for correspondence

Guoqian Jiang, MD, PhD at jiang.guoqian@mayo.edu.

Mapping HL7 CDA R2 Formatted Mass Screening Data to OpenEHR Archetypes

Shinji Kobayashi, Naoto Kume, Hiroyuki Yoshihara

The EHR department of Biomedical Informatics, Kyoto University, Kyoto, Japan

Abstract

Mass screening of adults was performed to manage employee healthcare. The screening service defined the data collection format as HL7 Clinical Document Architecture (CDA) R2. To capture mass screening data for nationwide electronic health records (her), we programmed a model within the CDA format and mapped the data items to the ISO13606/openEHR archetype for semantic interoperability.

Keywords:

Electronic Health Records, Semantics, Mass Screening

Introduction

Mass screening of adults has been performed for the early detection of diseases, and to limit healthcare expenditures by preventing lifestyle-related diseases in Japan. This screening service defines the data collection process as HL7 Clinical Document Architecture (CDA) R2 [1] and uses data for personal lifestyle prevention.

An electronic health record (EHR) project based on the ISO 13606/openEHR archetype was performed to use healthcare data for research purposes. To capture the mass screening data, we programmed a model within the CDA format and mapped the data to ISO13606/openEHR archetypes for semantic interoperability.

Methods

Datasets were initially analyzed to identify possible mass screening items. Subsequently, archetypes and templates were designed by the openEHR Archetype Editor and Ocean Template Designer, respectively. The archetypes were mainly derived from the openEHR Clinical Knowledge Manager with or without modifications. Since some concepts only concerned the usage of mass screening processes in Japan, they were newly designed to express such specialized concepts. We then mapped the HL7 CDA elements with the OID table to the path of elements in the openEHR template that was designed for capturing mass screening data.

Results

In accordance with mindmap analysis, mass screening data were recorded using an openEHR template, which was designed with 36 archetypes.

The main part of the CDA had a simple structure including item names and values. Each element was parameterized using items from a questionnaire and an OID table was mapped to concept models. For the mass screening CDA, doctors' evaluations of health were categorized as an OBS Reference

Information Model (RIM), but mapped as the Problem/Diagnosis archetype derived from the "EVALUATION" Reference Model (RM) within the openEHR archetype.

Discussion

To build EHR systems, standards are necessary to construct an interoperable data repository. Connected to the standardized data was a common problem, and we established mapping rules between the openEHR archetypes/templates and HL7 CDA. To efficiently and accurately establish mapping rules [2], terminology binding should be considered; however, no such terminology is available in Japan.

HL7 CDA-formatted data were transformed to openEHR archetypes/templates, thus establishing a possible method for building interoperable EHRs.

Conclusion

Mapping rules between HL7 CDA-formatted data and openEHR archetypes/templates was successfully performed in this study. For efficiency, such terminology should be widely adopted throughout Japan.

Acknowledgements

This research was supported by the "Clinical Study Oriented ICT Infrastructure Development Project - Sustainable Massive Health and Clinical Data Repository for Secondary Use" of the Japan Agency for Medical Research and Development.

References

- [1] R.H. Dolin, L. Alschuler, S. Boyer, C. Beebe, F.M. Behlen, and P.V. Biron, HL7 clinical document architecture, release 2, *Journal of the American Medical Informatics Association* **13** (2006), 30-39.

Address for correspondence

Shinji KOBAYASHI, MD, PhD

e-mail: kobayashi.shinji.5s@kyoto-u.ac.jp

Address: Room 506, 5th floor, Bldg. #9 Kyoto Research Park,
91 Awata-cho, Chudoji, Shimogyo-ku, Kyoto, Japan 600-8813

Implementation of an Archetype Data Set to Reuse Electronic Health Record Data in Clinical Decision Support Systems

Georgy KOPANITSA^{a,b}

^aInstitute Cybernetic Center, Tomsk Polytechnic University, Tomsk, Russia

^bTomsk State University for Architecture and Building, Tomsk, Russia

Abstract

The efficiency and acceptance of clinical decision support systems (CDSS) can increase if they reuse medical data captured during health care delivery. High heterogeneity of the existing legacy data formats has become the main barrier for the reuse of data.

Keywords:

Decision Support Systems, Clinical; Telemedicine; Electronic Health Records.

Introduction

Medical insurance companies perform quality assessment of medical treatment in Russia. The results of the quality control influence the payment that a clinic will get for the certain case of treatment. The quality assessment process is a well-formalized process performed by medical experts based on clinical standards and personal experience. However, different interpretation of clinical standards especially in regards of optional medical procedures add bias to the process. To avoid bias and to make the process transparent and independent the task of quality assessment can be delegated to a decision support system [1-6]. However, most of CDSSs still are stand-alone systems that require import-export routine with HISes. The goal of this paper is to present the development of an archetype based dataset for a CDSS-HIS data exchange.

Methods

Our study is based on the approach proposed by Marcos et al [7] where archetypes are used as a standardized mechanism for the interaction of a CDSS with an EHR. We used LinKEHR archetype editor to design archetypes. In our project, clinical data in EHR systems was stored in proprietary formats. All the HISes participating in the project provide routines that can import/export XML data with known XML schema. We have transformed this data into the openEHR standard format following the constraints defined in archetypes. To perform this transformation we used LinKEHR studio. Using LinKEHR studio we have defined mapping rules between an EHR data schema and CDSS archetypes. After the mapping has been completed, an XQuery script for each of the 10 HISes has been generated to transform the source data instance into an XML document compliant with the target archetypes.

Results

We have used archives of treatment cases from 10 clinics in Tomsk, Russia, participating in the pilot project. Each HIS database stores a complete electronic health record for each patient. On the first step, we have reviewed the openEHR archetype repository (Clinical Knowledge Manager (CKM) for the

archetypes that could be reused in our project. Although, the found archetypes had many of the necessary fields, there still was a need to design new archetypes and to add some of the fields useful for payment information and CDSS purposes like nonconformity, transfer summary and inference coefficients.

Conclusions

We have defined a set of archetypes to represent concepts for exchanging and analyzing interaction with insurance companies.

Acknowledgements

The research was supported by the Grant of a Russian President #AAAA-A16-116120810057-8. The research is carried out at Tomsk Polytechnic University within the framework of Tomsk Polytechnic University Competitiveness Enhancement Program grant.

References

- [1] J.L. Cardoso de Moraes, W.L. de Souza, L.F. Pires, A.F. do Prado, A methodology based on openEHR archetypes and software agents for developing e-health applications reusing legacy systems, *Comput Methods Programs Biomed* **134** (2016):267-287.
- [2] H.J. Kam, J.A. Kim, I. Cho, Y. Kim, R.W. Park, Integration of heterogeneous clinical decision support systems and their knowledge sets: feasibility study with Drug-Drug Interaction alerts, In: *AMIA Annu Symp Proc* **2011** (2011).
- [3] G. Kopanitsa, Evaluation study for an ISO 13606 archetype based medical data visualization method, *J Med Syst* **39** (2015):82.
- [4] G. Kopanitsa, A qualitative study of the barriers and opportunities for adoption of web-portals for doctors and patients in Russia, *J Med Syst* **41** (2017):62.
- [5] G. Kopanitsa, A. Karpov, G. Lakovenko, A. Laskovenko, Implementation of a web portal for diabetes patients using open source data visualization libraries, *Stud Health Technol Inform* **224** (2016):189-194.
- [6] G. Kopanitsa, V. Yampolskiy, Approach to extract billing data from medical documentation in Russia - lessons learned, *Stud Health Technol Inform* **210** (2015).
- [7] F. Pecoraro, D. Luzzi, F.L. Ricci, Data warehouse design from HL7 clinical document architecture schema, *Stud Health Technol Inform* **213** (2015):139-142.

Address for correspondence

TPU, 634050 Lenina 30, Tomsk, Russia,
georgy.kopanitsa@gmail.com

Correcting Ontology Errors Simplifies Visual Complexity

Hao Liu, Ling Zheng, Yehoshua Perl, Yan Chen, Gai Elhanan

Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102-1982 USA

Abstract

In previous research we have shown that hierarchically complex overlapping concepts have a higher error rate of errors versus control concepts. In this poster we show an example from Neoplasm concepts of the NCI thesaurus (NCIt) demonstrating that erroneous overlapping concepts, reflected in the partial-area units of a partial-area taxonomy, display visual complexity. Furthermore, correcting these erroneous concepts causes visual simplification.

Keywords:

Ontology Quality Assurance; Ontology Visualization.

Introduction

In a current study of Quality Assurance of Neoplasm concepts of the NCIt, we have shown that hierarchically complex concepts have higher error rate than control concepts. Those complex concepts are called overlapping concepts, since they belong simultaneously to multiple units (groups) called partial-area of similar concepts in a partial-area taxonomy [1] (taxonomy for short) of the NCIt *Neoplasm* subhierarchy. Such overlapping concepts are semantically complex due to the compound semantics originated from different partial-areas. For formal definition see [2]. In this poster, we bring an example which illustrates a correlation between erroneous overlapping concepts and their visual complexity. To show this we demonstrate that the impact of correcting the errors of these overlapping concepts brings to the visual simplification of the complexity reflected for the erroneous concept.

Methods

For visual reflection of the complexity before and after the corrections, we utilize *Venn Diagrams*. Figure 1(a) displays an excerpt of six concepts of *Neoplasm* subhierarchy structure in the area with three roles {*Disease Has Abnormal Cell*, *Disease Has Associated Anatomic Site*, *Disease Has Finding*}. That is, each of the six concepts has exactly these roles. The concepts of each unit (partial-area) are enclosed in a bubble of a different color. Visual complexity is reflected by the existence of concepts which belongs simultaneously to several bubbles. For example in Figure 1(a), the two concepts *Recurrent Childhood Central Nervous System Neoplasm* and *Recurrent Childhood Spinal Cord Neoplasm* are overlapping concepts in the audited sample. They belong to both the red partial-area rooted at *Childhood Central Nervous System Neoplasm* and the green partial-area rooted at *Recurrent Central Nervous System Neoplasm*.

Furthermore, when the two overlapping concepts are corrected by adding the role *Disease Has Primary Anatomic Site*, the modeling changes. When looking at the same concepts after remodeling, the view through the prism of a taxonomy reflects a simpler visual presentation. In particular these two concepts are no longer overlapping concepts of the two partial-areas in Figure 1(a). After the remodeling, they appear in a separate

area with four roles {*Disease Has Abnormal Cell*, *Disease Has Associated Anatomic Site*, *Disease Has Finding*, *Disease Has Primary Anatomic Site*}. A new blue bubble (partial-area) rooted at the concept *Recurrent Childhood Central Nervous System Neoplasm*, contains the two corrected concepts.

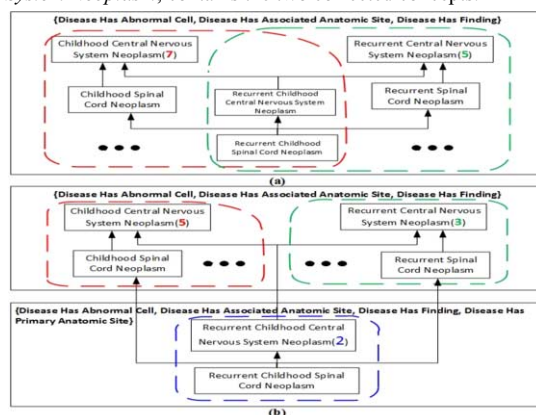


Figure 1 - (a) before correction of two overlapping concepts y adding a new role; (b) after

Results

What we see in Figure 1, is that in contrast to the visual complexity in Figure 1(a) reflecting errors in modeling, the corresponding visual complexity of the partial-areas of the six concepts, following the remodeling, is simpler. That is, corrections of the modeling errors in overlapping concepts leads to visual simplification, when reviewing the taxonomy compact representation of the ontology. A future study is planned to confirm the following hypothesis: The visual presentation of erroneous overlapping concepts in a taxonomy, is simplified when the errors are corrected.

The future study will investigate in what percentage of erroneous overlapping concepts, a visual simplification is obtained by correcting the errors.

Conclusion

In this poster we demonstrate how erroneous concepts are manifested by visual complexity. Correcting them leads to visual simplification.

References

- [1] Y. Wang, M. Halper and et al, *Auditing complex concepts of SNOMED using a refined hierarchical abstraction network*, *J Biomed Inform* **45**(1) (2012), 1-14.
- [2] Y. Wang, M. Halper, H. Min, Y. Perl, Y. Chen, and K.A. Spackman, *Structural methodologies for auditing SNOMED*, *J Biomed Inform* **40** (2007), 561-581.

Using Semantic Technologies to Extract Highlights from Care Notes

Vanessa Lopez^a, Grace McCarthy^a, Joao Bettencourt-Silva^a, Marco Sbodio^a, Natasha Mulligan^a, Fabrizio Cucci^a, Stephane Deparis^a, Ciaran Hennessy^b, Nagesh Yadav^b, Karie Kelly^b, Russell Olsen^b, Conor Cullen^b, Spyros Kotoulas^a

^aIBM Research Ireland, Dublin, Ireland, ^bIBM Watson Health, Dublin, Ireland

Abstract

We propose a cognitive system for patient-centric care that leverages and combines natural language processing, semantics, and learning from users over time to support care professionals working with large volumes of patient notes. The proposed methods highlight the entities embedded in the unstructured data to provide a holistic semantic view of an individual. A user-based evaluation is presented, showing consensus between the users and the system.

Keywords:

Health Information Systems, Artificial Intelligence

Introduction

The adoption of electronic health records have contributed to a growing volume of data while promising to improve quality of care and reduce costs. However, in a system of record, insights about a patient with multiple clinical and social issues are often scattered among several notes. New methods are needed to provide automated extraction of highlights and insights to relevant information to care practitioners [2]. Our system captures knowledge from care professionals' observations, often in an unstructured form, to create a holistic patient-centred view, consisting of entities with explicit semantics extracted from notes. Care professionals are then presented with highlights on the most relevant entities, helping them making informed and personalised decisions.

Methods

The system first annotates raw case notes associated to patients to extract entities. The annotation component consolidates the annotations from different text annotators. Currently, we use Advanced Concept Insights (ACI), a clinical IBM annotator based on ICD-10 terminology and AlchemyAPI to annotate concepts not included in ICD-10 – mainly social determinants. Second, the terminology service maps the annotations from the text to semantic entities in well-defined ontology models. Third, the note highlights component learns to rank over time the relevant determinants of health for a patient based on users' feedback. A key challenge is to deal with the diversity of the social and clinical domains. The terminology service is built by creating a Lucene index over: (1) the ICD-10 clinical hierarchy; (2) a Linked Data subset of Freebase and DBpedia, covering the following semantic types: clinical (procedures, medication, symptoms, etc), care services and social determinants (risk factors, activities of daily living, etc). The added value behind using a terminology service is that: (1) having a URI gives a global meaning to abstract from the annotators heterogeneous

models of choice, e.g., representing synonyms like Eye-drops and Ocular lubricant with the same entity; (2) a semantic type is used to filter the entities that are most likely to be relevant for the user and for organizing them into views.

Results

We created a gold standard to measure whether our system can extract similar entities from notes as to what users will choose. We selected 20 patient cases based on social and clinical collections of notes [1] and real care workers' notes anonymized and de-identified. To create the relevance judgments, 15 evaluators were asked to highlight all relevant keywords (user annotations) and select the top 10 highlights for a case. On average each case had 91 entities and was evaluated by 3 evaluators to measure agreement: in 80% of the cases, more than two evaluators highlighted an entity. We measured an F1 score of 0.67 (i.e., the system ability to find all relevant and only relevant annotations), with 0.86 Precision and 0.56 Recall for all the entities extracted by the system with respect to all user annotations with at least moderate agreement. If we only consider the entities with strong agreement and the top-n annotations, R (coverage) increases up to 0.78, indicating that with enough learning, the performance can improve for the entities that matter the most. While 25% of entities were not assigned a semantic type due to the lack of models' coverage, if only a clinical annotator is used (i.e., ACI) P increases to 0.94 but R drops to 0.27, showing the importance of including social concepts.

Conclusion

The system targets an interdisciplinary care team. While social aspects are not as well defined as clinical, we show promising results to provide a semantic view on top of patient notes, based on a shared terminology that reuses well-known heterogeneous ontologies. In the future we expect to determine whether notes highlights offer a better experience for care professionals than simply reading case notes.

References

- [1] <http://mtsamples.com/>
- [2] Pivovarov, R., Elhadad, N., Automated methods for the summarization of electronic health records. J Am Med Inform Assoc 22(5), (2015), 938-47.

Address for correspondence

Vanessa Lopez, vanlopez@ie.ibm.com. IBM Research Ireland.

Challenges in Archetypes Terminology Binding Using SNOMED-CT Compositional Grammar: The Norwegian Patient Summary Case

Luis Marco-Ruiz, Rune Pedersen

Norwegian Centre for e-Health Research, University Hospital of Northern Norway, Tromsø, Norway

Abstract

In order to cover the requirements for interoperability in the Norwegian context, we studied the terminology binding of archetypes to terminology expressions created with the SNOMED-CT compositional grammar. As a result we identified important challenges categorized as technical, expressivity, human, and models mismatch.

Keywords:

Systematized Nomenclature of Medicine; Vocabulary, Controlled; Terminology.

Introduction

In order to fulfill the interoperability requirements of the Norwegian national health IT strategy in a scalable manner, machine-interpretable representation of an archetype's clinical semantics is needed. As a result, we studied the feasibility of using SNOMED-CT compositional grammar to build semantic models that distilled the implicit ontology contained in archetypes [1] (i.e. leaving data constraints aside).

Methods

In collaboration with the National Editorial Group of Archetypes (NRUA) we selected the most representative archetypes of the Norwegian patient summary. For each archetype, terminology binding was attempted by creating a projection of its clinical semantics using the SNOMED-CT compositional grammar. When some element/section of the archetype could not be represented using a SNOMED-CT expression, we defined a new expression and tagged the cause. Additionally, when an impediment caused by the complexity of the process, lack of tooling etc. was found, we tagged it. Afterwards we reviewed all the tags and classified them into categories of challenges.

Results

Four types of challenges were identified:

Technical challenges were related to: a) lack of support in archetypes to include verbose post-coordinated expressions in their term bindings section; b) lack of tooling to assist the binding process suggesting valid concepts/attributes while building the terminology expression.

Expressivity challenges were related to: a) variation of the original meaning of the archetype element introduced by the terminology concept; b) lack of expressiveness of SNOMED-CT context model; c) lack of candidates in SNOMED-CT to represent some archetype elements.

Human challenges were related to: a) doubts in determining which sections of the archetype should be represented in the SNOMED-CT expression; b) selecting the appropriate terminology hierarchy (e.g. whether to represent blood pressure with the concept from Observable Entity or Clinical Finding hierarchies).

Models mismatch challenges were related to: a) alignments of the archetype contextual information and the SNOMED-CT context model; b) low coverage in SNOMED-CT for the attributes in the protocol section; c) overlap of semantics in the archetype (the archetype Problem_Diagnosis corresponds to two concepts from different hierarchies in SNOMED-CT).

Discussion

The results show that major challenges are present when representing an archetype's clinical semantics as SNOMED-CT expressions. First, technical challenges show that clinical modelers need to clarify if post-coordinated expressions should be placed within the archetype definition. An alternative is to let terminology specialists maintain complex expressions in external repositories and use Linked Data principles to reference them [2]. Second, expressivity challenges show that better guidelines elicited by archetypes and SNOMED-CT editors are needed to determine what elements can be bound to SNOMED-CT expressions, and what elements should be bound to other domain ontologies (if needed). Third, technical and human challenges show that defining and binding SNOMED-CT expressions requires the development of support tools. However, further investment in these tools [3] is needed to allow them supporting such functionality.

References

- [1] Goossen W, Goossen-Baremans A, van der Zel M. Detailed Clinical Models: A Review. *Healthcare Informatics Research* 2010;16:201. doi:10.4258/hir.2010.16.4.201.
- [2] Marco-Ruiz L, Pedrinaci C, Maldonado JA, Panziera L, Chen R, Bellika JG. Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach. *Journal of Biomedical Informatics* 2016;62:243–64. doi:10.1016/j.jbi.2016.07.011.
- [3] Sundvall E, Qamar R, Nystrom M, Forss M, Petersson H, Karlsson D, et al. Integration of tools for binding archetypes to SNOMED CT. *BMC Med Inform Decis Mak* 2008;8 Suppl 1:S7. doi:10.1186/1472-6947-8-S1-S7.

Address for correspondence

Norwegian Centre for e-Health Research, University hospital of Northern Norway, P.O. Box 35, N-9038 Tromsø, Norway.

E-mail: Luis.Marco.Ruiz@chealthresearch.no

Terminology Status APIs – Mapping Obsolete Codes to Current RxNorm, SNOMED CT, and LOINC Concepts

Lee Peters, Thang Nguyen and Olivier Bodenreider

U.S. National Library of Medicine, Bethesda, MD, USA

Abstract

We created the Terminology Status Application Programming Interface (API) to assist users in mapping obsolete codes to current RxNorm, SNOMED CT and LOINC concepts. Use cases include support for information retrieval, maintenance of value sets, and analytics of legacy clinical databases. Our terminology status APIs typically receive over 4 million calls per month on average.

Keywords:

RxNorm; Systemized Nomenclature of Medicine; Logical Observation Identifiers Names and Codes

Motivation

In medical terminologies, new releases introduce new concepts, remove obsolete concepts and reorganize others through remapping. Managing change in large terminologies can be burdensome to users. We created the Terminology Status Application Programming Interface (API) to assist users in mapping obsolete codes to current RxNorm, SNOMED CT and LOINC concepts.

Terminology Status APIs

The *status* APIs have been developed to help manage change in three major medical terminologies, whose use is required for the Meaningful Use certification criteria: RxNorm, SNOMED CT and LOINC. While the RxNorm service has been part of the RxNorm API for several years, we developed equivalent services for SNOMED CT (including concepts specific to the U.S. extension of SNOMED CT) and LOINC more recently, for the purpose of validating value sets for clinical quality measures.

The **RxNorm status API** [1] returns three elements: *Status* – active, “alien” (i.e., present in one of the source terminologies, but without an RxNorm type), remapped, retired or unknown; *Last active release date*; and *Concept information* for active and remapped status.

The **SNOMED CT status API** [2] returns three elements: *Current status* – retired, active or moved; *Effective time* – the version of the latest change; and *SNOMED CT identifiers* for remapped concepts.

The **LOINC mapto API** [3] returns an array of LOINC identifiers when the original identifier is remapped.

All three APIs are available in two flavors, SOAP-based and RESTful. Each API also has a version function so that users can determine which version of the data set is being used. The services are updated as new terminology versions become available. Figure 1 shows an example of query to the SNOMED CT status API (RESTful version), along with the information returned.

```
https://rxnav.nlm.nih.gov/REST/SnomedCT/status?id=187959005

<snomedctStatus>
  <inputId>187959005</inputId>
  <status>2</status>
  <effectiveTime>20120131</effectiveTime>
  <mappedId>449627008</mappedId>
  <mappedId>449628003</mappedId>
</snomedctStatus>
```

Figure 1 - SNOMED CT Status example. Status 2 indicates that the input concept (Malignant neoplasm of long bones of leg) was replaced by the 2 concepts, Malignant neoplasm of long bone of lower limb and Malignant neoplasm of long bone of lower leg

Use Cases

Use cases for these *status* APIs include: 1) Supporting information retrieval systems that accept concept identifiers as queries (e.g., MedlinePlusConnect uses this service); 2) Determining outdated or remapped drug concepts in clinical drug databases (e.g., updating value sets); and 3) Supporting analytics of older datasets coded to past versions of the terminologies.

Our terminology status APIs typically receive over 4 million calls per month on average.

Acknowledgements

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

References

- [1] United States National Library of Medicine (2017), RxNorm API, <https://rxnav.nlm.nih.gov/RxNormAPIs.html>
- [2] United States National Library of Medicine (2017), Snomed CT Status API, <https://rxnav.nlm.nih.gov/SnomedCTAPI.html>
- [3] United States National Library of Medicine (2017), LOINC Mapto API, <https://rxnav.nlm.nih.gov/LoincAPI.html>

Address for correspondence:

RXNAVINFO@LIST.NIH.GOV

From a Categorical Structure to a Concept Model: The International Classification of Health Interventions (ICHI)

Jean-Marie Rodrigues^{a,b,f}, Sukil Kim^{c,d}, Syed Aljunid^e,
Jae Jin Lee^e, Huib Ten Napel^g, Béatrice Trombert^{a,b}

^a INSERM U1142, LIMICS, Paris, France

^b Department of Public Health and Medical Informatics, Université Jean Monnet of Saint Etienne, France

^c The Catholic University of Korea, Seoul Korea

^d WHO Collaborating Center for International Classifications and Terminology, Seoul, Korea

^e International Centre for Casemix and Clinical Coding, National University of Malaysia

^f WHO Collaborating Center for International Classifications in French Language, Paris, France

^g WHO Collaborating Center for International Classifications in the Netherlands

Abstract

The International Classification of Health Interventions (ICHI) alpha2 2016 Section 1 Interventions on Body Systems and Functions is based on ISO 1828 international standard named categorical Structure (CAST). This is not sufficient to represent the meaning of ICD9-CM Volume 3 labels. We propose to modify it by using the SNOMED CT concept model.

Keywords:

Vocabulary, controlled, Systematized nomenclature of Medicine

Introduction

We propose to apply ontology methods [1-3] for building a new generation international classification of procedures.

Methods

305 ICD9-CM volume 3 codes mapped many to one out of 113 The International Classification of Health Interventions (ICHI) codes. ICHI coding system displays the following:

- PZX DB AF means PZX Any site DB Application of Pharmaceuticals including chemotherapy AF Percutaneous transluminal /Transparietal intraluminal access ICD-9CM volume 3 mapped codes are different for implantation (00.10), infusion (00.11), injection (00.13), immunization (99.12). We compare ICD9 CM codes titles with SNOMED CT (SCT) procedure concept model to identify required modifications.

Results

The following modifications are proposed:

- The ICHI axis Action should be duplicate to express the intent and the deed.
- The ICHI Target axis should be extended to pathology as calculus, and to medical devices as pacemaker.
- The ICHI Target axis should be duplicate in Direct Target and Indirect Target.
- The ICHI Means axis should be extended to medical devices and drugs.

Conclusion

The analysis of the SCT concept model provides a method to increase ICHI semantic interoperability.

Acknowledgments

The WHO FIC network.

References

- [1] B. Smith, Applied ontology: A new discipline is born, *Philosophy Today* **29** (1998), 5-6.
- [2] F. Baader F, *The Description Logic Handbook: Theory, Implementation, and Applications, 2nd Edition*, Cambridge University Press, 2007.
- [3] SNOMED CT Document Library. Available from: <https://confluence.ihtsdotools.org/display/DOC>

Address for correspondence

RODRIGUES Jean-Marie

Email: rodriguesjeanmarie2@gmail.com

POLE.VAULT: A Semantic Framework for Health Policy Evaluation and Logical Testing

Arash Shaban-Nejad^a, Anya Okhmatovskaia^b, Eun Kyong Shin^a, Robert L. Davis^a, David L. Buckeridge^b

^a UTHSC-Oak Ridge National Lab (ORNL) Center for Biomedical Informatics,

Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN, USA

^b McGill Clinical & Health Informatics, Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada

Abstract

The major goal of our study is to provide an automatic evaluation framework that aligns the results generated through semantic reasoning with the best available evidence regarding effective interventions to support the logical evaluation of public health policies. To this end, we have designed the POLicy EVALuation & Logical Testing (POLE.VAULT) Framework to assist different stakeholders and decision-makers in making informed decisions about different health-related interventions, programs and ultimately policies, based on the contextual knowledge and the best available evidence at both individual and aggregate levels.

Keywords:

Public Health Systems Research, Program Evaluation, Evidence-Based Practice

Introduction

Root cause analysis (RCA) [1] is a technique for detecting and describing the causality path for a problem and recommending remedial actions. Using contextual knowledge captured in ontologies, we analyze public health programs and policies and discuss how context influences outcomes. To evaluate public health programs, we begin by studying their impacts on target populations/communities and the local and global environment (e.g. changed public knowledge, a situation or behavior). The stakeholders (i.e. public health institutions, policy-makers and advocacy groups, practitioners, insurers, clinicians, and researchers) often need to consider several factors and collect information from multiple resources to answer questions regarding the effectiveness of public health interventions. The POLicy EVALuation & Logical Testing (POLE.VAULT) Framework aims to facilitate knowledge-based evaluation of public health interventions and policies.

Methods and Results

POLE.VAULT evaluates public health interventions by assessing their logic models that represent the logical relationships between resources, activities, outputs, outcomes and impacts of a public health program. This is important because, regardless of whether evaluations are process- or outcome-oriented, or qualitative or quantitative, they need to assess whether critical program components or activities were implemented and whether they had an impact on mediating outcomes, important behaviors and overall health goals.

To evaluate different sections and components of logic models, we developed a set of semantic queries that examine whether a program/intervention had the anticipated impact(s) in the population health status (positively, negatively, or unchanged). Then we align the logic model with the semantic knowledge platform and the conceptual model created using best available evidence in the domain.

For example, using PopHR [2], a knowledge-based population health record, we deploy evidence that captures knowledge about the burden of a disease due to specific causes (causal, correlational, and explanatory knowledge) and then combine it with evidence related to guidelines and protocols regarding the effectiveness of interventions and policies, confronting those causes, along with information about the target populations and their environments.

Discussion and Conclusion

The POLE.VAULT platform allows assessing public health programs/interventions and their outcomes through a set of individual and group level indicators. It also intends to assist in the comparative analysis of effectiveness and cost of different public health interventions and policies. Moreover, in the future our platform intends to evaluate programs based on their coverage and effectiveness by highlighting the proportion of the defined objectives and goals that a program is capable of addressing, within a specific resource range (e.g. time period, budget bracket and human resources).

References

- [1] T.O.A. Lehtinen, M.V. Mäntylä, J. Vanhanen, Development and evaluation of a lightweight root cause analysis method (ARCA method) – Field studies at four software companies, *Information and Software Technology* 53(10) (2011), 1045–1061.
- [2] A. Shaban-Nejad, M. Lavigne, A. Okhmatovskaia, D.L. Buckeridge, PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data, *Ann N Y Acad Sci* 1387 (1) (2017), 44–53.

Address for correspondence

Arash Shaban-Nejad, email: ashabann@uthsc.edu, UTHSC-ORNL Center for Biomedical Informatics, 497R, Le Bonheur Research Center 50 N Dunlap St., Memphis TN 381032893.

Clinical Workflow Modeling in Obstetrics: Hepatitis B in Pregnancy

Fadi Shamoon^{a,b}, Harald Leitch^c, Jeroen S. de Bruin^{b,d},
Andrea Rappelsberger^d, Klaus-Peter Adlassnig^{b,d}

^a University of Applied Sciences FH Campus Wien, Favoritenstraße 226, A-1100 Vienna, Austria

^b Medexter Healthcare GmbH, Borschkegasse 7/5, A-1090 Vienna, Austria

^c Department of Obstetrics and Gynecology, Medical University of Vienna, Währinger Gürtel 18–20, A-1090 Vienna, Austria

^d Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria

Abstract

Evidence-based clinical guidelines positively effect physician decision-making. Actionable clinical guidelines that actively trigger alerts, reminders, and instructive texts will increase effectiveness. We applied Activiti, a Business Process Model and Notation language system to model a clinical guideline for the prevention of mother-to-child transmission of hepatitis B as a computerized clinical workflow. Furthermore, we implemented an interconnected Arden-Syntax-based medical rule engine, which is part of the ARDENSUITE software.

Keywords:

Practice Guidelines; Decision Support Systems, Clinical; Pregnancy Complications.

Introduction

This study addresses the (partial) implementation of the “Hepatitis B in pregnancy” guideline: a clinical workflow providing evidence-based instructions on preventing mother-to-child transmission of hepatitis B during delivery. The guideline is in use at the Department of Obstetrics and Gynecology at Vienna General Hospital, Austria (Figure 1).

Methods

Business Process Model and Notation: the open-source workflow platform Activiti for the development of Business Process Model and Notation (BPMN) workflows [1, 2].

Arden Syntax: data access and knowledge-based clinical guidelines were implemented in HL7’s standard notation Arden Syntax part of the ARDENSUITE solution [3-5].

Results

BPMN translation permitted automation or the provision of reminders to the pediatrician of the following tasks:

- Active and passive immunization after childbirth
- Specific breastfeeding recommendation depending on the results of HBsAg and hepatitis B PCR testing
- Availability of order forms for in-house or external lab testing for HBsAg or HBsAg and/or hepatitis B

Conclusion

We presented a clinical workflow implementation connecting

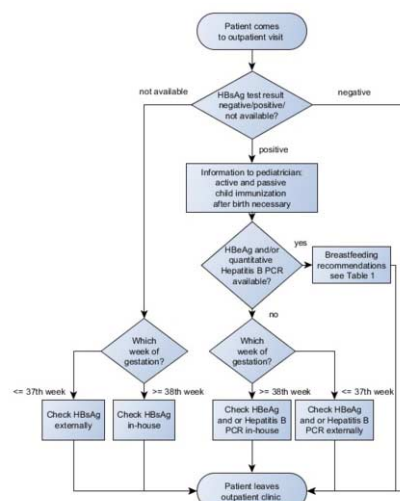


Figure 1 - Hepatitis B clinical guideline

a BPMN-based workflow using the ARDENSUITE. This simplifies complex clinical guidelines by dividing them into workflow, knowledge processing, and data access steps and separates patient data access and knowledge-based decisions from the actual workflow/workflow decisions. This is essential for integrating evidence-based clinical guidelines into the patient care process.

References

- [1] Activiti, 2016. Available from: <http://activiti.org/components.html>
- [2] T. Rademakers, *Activiti in Action – Executable Business Processes in BPMN 2.0*, Manning Publications, Shelter Island, 2012.
- [3] ARDENSUITE, 2016. Available from: <http://www.medexter.com/products/ardensuite>
- [4] R.T. Fielding, *Architectural Styles and the Design of Network-Based Software Architectures*, Dissertation, University of California, Irvine, 2000.
- [5] K.P. Adlassnig, K. Fehre, Service-Oriented Fuzzy-Arden-Syntax-Based Clinical Decision Support, *Indian Journal of Medical Informatics* 8 (2014), 75-79.

Address for correspondence

Dr. Klaus-Peter Adlassnig, FACMI

E-mail 1: klaus-peter.adlassnig@meduniwien.ac.at;

E-mail 2: kpa@medexter.com

CoMetaR: A Collaborative Metadata Repository for Biomedical Research Networks

Mark R. Stöhr, Gudrun Helm, Raphael W. Majeed, Andreas Günther

UGMLC, German Center for Lung Research (DZL), Justus-Liebig-University, Giessen, Germany

Abstract

The German Center for Lung Research (DZL) is a research network with the aim of researching respiratory diseases. To perform consortium-wide queries through one single interface, it requires a uniform conceptual structure. No single terminology covers all our concepts. To achieve a broadly accepted and complete ontology, we developed a platform for collaborative metadata management “CoMetaR”. Anyone can browse and discuss the ontology while editing can be performed by authenticated users.

Keywords:

Metadata; Terminology; Intersectoral Collaboration

Introduction

The DZL is a consortium of multiple lung research institutions. We collectively pursue the goal to find ways of preventing and curing respiratory diseases. There are many different local data storing systems in use, e.g. CentraXX, Filemaker etc. This circumstance hinders researchers from performing consortium-wide queries quickly and with least effort. Therefore, a central data warehouse (i2b2) is used to which every site uploads their data. Semantic integration of lung research data requires not just one but multiple existing terminologies like LOINC and SNOMED-CT, in addition to custom lung research specific concepts. The OBO Foundry lists four principles for developing a new ontology [1]: (i) be developed in a collaborative effort, (ii) use common relations that are unambiguously defined, (iii) provide procedures for user feedback and for identifying successive versions and (iv) have a clearly bounded subject-matter. Aim of this project is to realize a platform that visualizes the DZL metadata ontology and enables medical documentalists and researchers from all participating institutions to take part in the development process.

Methods

Our requirement analysis resulted in the following statements: The ontology has to be visualized and searchable through an user interface, which is accessible by any person without additional software. Medical documentalists need to maintain the ontology, which should be possible without additional software. The description format has to be extendable. Any DZL member should be able to contribute expertise. Every term of the ontology needs to be discussable and discussion history itself accessible. Trained specialists from anywhere should be able to take part in editing the ontology independently and simultaneously. Changes to the ontology need verification, e.g. every concept needs to be labeled and properly integrated into the hierarchy. Investigation of existing solutions [2,3,4] did not lead to satisfying results.

Results

We developed a platform for independently maintaining an ontology from different sites combined with a web interface for visualization and concept related discussion. The web interface is based on standard technologies in order to guarantee accessibility. All metadata concepts can either be explored through an expandable tree or a search form. For each concept the user is offered details and a discussion board. RDF was chosen as ontology description format because of it's basic purpose for graph description, it's extensibility as well as the possibility of editing in simple text editor. Concept hierarchy is realized through the Simple Knowledge Organization System (SKOS) relations “broader” and “narrower”. All RDF files are stored in a GIT repository, to which any person with access can upload data. Transmissions are verified through syntactical and semantical tests. Afterwards, the updated ontology is immediately loaded into the i2b2 server and a triple-store. The latter provides all information for the web interface via AJAX and standard SPARQL queries.

Discussion

During development we focused on using existing standards. CoMetaR may be used for arbitrary SKOS ontologies. It is especially suitable for management of biomedical and other evolving ontologies. Additional functionality may be integrated into the web interface through javascript extensions.

Conclusions

We developed a software for collaborative management of metadata ontologies. Specialists from all participating institutions are able to view and contribute. All source code is open source and available at <https://github.com/dzl-dm/cometar>.

References

- [1] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall et al., The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration., *Nat Biotechnol* **25** (2007), 1251-1255.
- [2] Protégé, <http://protege.stanford.edu>, Accessed 4/17.
- [3] BioPortal, <http://bioportal.bioontology.org>, Accessed 4/17.
- [4] OBO-Edit, <http://oboedit.org/>, Accessed 4/17.

Address for correspondence

Mark R. Stöhr, e-mail: mark.stoechr@innere.med.uni-giessen.de

Using an Ontology-Based Approach to Handle Author Affiliations in a Large Biomedical Citation Database

Haixia Sun^a, Junlian Li^a, Yingjie Wu^a, Lei Wang^a, Kin Wah Fung^b

^a Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China

^b National Library of Medicine, Bethesda, MD, USA

Abstract

To handle differences in affiliation names submitted with biomedical journal articles, we build an affiliation knowledge base named Authority File for Affiliations (AFA) based on ontology principles. There are currently 113,700 affiliation concepts with about 583,700 affiliation names. The AFA becomes an essential tool in managing citation information and data analysis.

Keywords:

Databases, Organizational Affiliation, Information Management

Introduction

CBM (Chinese Biomedical Database) contains more than 9 million abstracts of journal articles since 1978, covering over 1,800 Chinese biomedical journals. We build an affiliation ontology (Authority File for Affiliations, AFA) to manage author affiliation information. Previous work on affiliation information management mostly focused on organization names and synonyms, such as the work by Jonnalagadda et al. [1]. Our conceptualization of the AFA enables us to represent attributes of affiliations, their evolution (e.g., splits and merges) and relationships (e.g., subsidiary, associated hospital).

Methods

We developed a web-based application using Java and SQL database to support AFA editing. Original affiliation names were extracted from citations, then cleaned and normalized. Each affiliation concept contained three kinds of names: official name, preferred name and synonyms. The preferred name was the most frequently used name in the citations. Attributes were added to the affiliation concepts, such as organization type, location and specialty. Relationships between affiliations included division_of, school_of, renamed_from, merged_from, and affiliated_hospital_of etc. Affiliations were linked to their citations through their unique identifiers (UI).

Results

There are currently 113,700 affiliation concepts with 583,700 affiliation names, originating from more than 2 million unique original affiliation strings and 7 million abstracts. Each affiliation has one or more organization types.

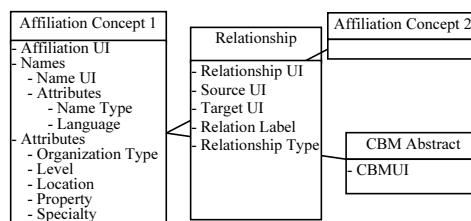


Figure 1- Ontology-based Data Model for AFA

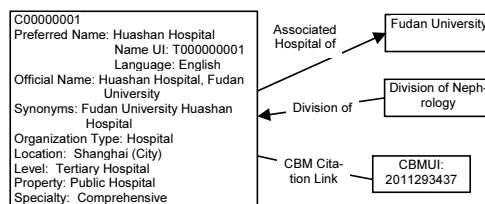


Figure 2- Example of "Huashan Hospital" In AFA

Discussion

We anticipate that AFA will greatly enhance our ability to handle variations and resolve ambiguity in affiliation names related to different styles of usage, non-standard abbreviations and occasional typos. It facilitates searching and retrieval by providing a rich set of synonyms. The explicit representation of inter-affiliation relationships and attributes also enables us to perform analysis at different levels of granularity (e.g., at the university, school or department level), and according to certain properties of the affiliations (e.g., organizational type).

Conclusion

An ontology-based approach is an efficient way to handle lexical variations in affiliation information and will facilitate data retrieval and analysis.

Acknowledgements

This paper is funded by the Fundamental Research Funds for the Central Universities (Grant No. 2016RC330006).

References

- [1] Siddhartha Jonnalagadda, Philio Topham. NEMO: Extraction and normalization of organization names from PubMed affiliation string. *Journal of Biomedical Discovery and Collaboration* (2010)5:50-75.

Address for correspondence

Haixia Sun, Email: sun.haixia@imicams.ac.cn

Self-Reported eHealth Literacy Among Undergraduate Nursing Students in Selected Districts of Sri Lanka

Shaluni Tissera^a, Nishan Silva^b

^a School of Nursing, International Institute of Health Sciences, Welisara, Sri Lanka

^b Academic Department, International Institute of Health Sciences, Welisara, Sri Lanka

Abstract

Nurses as the largest Health Care Workers group, are extremely important in promoting eHealth. Before promoting eHealth in a structured system, personal use of eHealth like as for Health Information Seeking; is important. Therefore study was done on the use of electronic health information sources among the Sri Lankan Nursing students. It showed that though they do use a wide range of sources, but they are unable to use them specifically and reliably.

Keywords: .

Telemedicine, Health Literacy, Computers

Introduction

The Ministry of Health in Sri Lanka is leading several initiatives to digitalize the National health system. These include familiarizing both Health care workers and the public with the electronic health and using electronic sources to access health information. However, the current levels of eHealth literacy and the related practices of the Sri Lankan nurses, the largest category of professional health care workers, is not known [1]. So the objectives of this study were to assess the eHealth Literacy and among 3rd year Nursing Students and its relationship with their other IT practices.

Methods

A descriptive cross sectional study was conducted among 136 conveniently selected 3rd year students from few selected Nurses training schools in Sri Lanka (Matara from the southern, Ampara from the eastern and Kandy from the central provinces) in 2016, using the eHEALS self reported eHealth Literacy Scale [2].

Results

88 (64.7 %) uses only their mobile phones to access internet and 20.6 % uses multiple devices at multiple locations. 27.9 % (32) of the nursing students used internet only to access social media. Only 16.2 % used internet for their daily routine work. Cronbach's alpha for the scale was 0.845. Mean eHEALS score (min 8; max 40) was 30.8 with a standard deviation of 3.479. The mean scores for individual scale items were most for knowing where to find helpful health resources (3.96 out of 5) and least for being able to discriminate between high and low quality resources (3.63). The mean eHEALS score does not depend on the duration of internet use ($p=0.532$).

Conclusion

The eHealth literacy among 3rd year Nursing students was fairly satisfactory. But the ability to use electronic resources reliably seems to be questionable.

Acknowledgements

The Authors would like to acknowledge the support of the Administration of all the visited Nursing Schools.

References

- [1] R.G. Booth, Educating the future eHealth professional nurse, *International Journal of Nursing Education Scholarship* 3 (2006), 1187.
- [2] R. Van der Vaart, A.J. van Deursen, C.H. Drossaert, E. Taal, J.A. van Dijk, and M.A. van de Laar, Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations, *Journal of medical Internet research* 13 (2011), e86.

Address for correspondence

Shaluni Tissera, shalunitissera@gmail.com, +94773539288

The Analysis of Medical Adverse Events Related to Electronic Health Records in Nursing Services

Mika Tomita^a, Natsumi Kishi^b, Mariko Iwasawa^c

^a Department of Nursing, School of Health Sciences, Ibaraki Prefectural University of Health Sciences, Ibaraki, Japan,

^b Department of Nursing, Jichi Medical University Hospital, Tochigi, Japan,

^c Faculty of Library, Information and Media Science, University of Tsukuba, Ibaraki, Japan

Abstract

This pilot study investigated problems of electronic health records (EHR), which have been used by nurses as a clinical decision tool. The investigation was conducted based on case records retrieved from the national database of medical adverse events. Detailed data related to nursing services must continue to be collected to establish a clearer linkage of EHR data and scholarly information.

Keywords:

Nurses, computerized medical records systems, safety management.

Introduction

Electronic health records (EHR) contribute to the collection of a large amount of patient data and thus are useful to provide effective nursing care. Data collected from the EHR can also be used to perform high-quality safety management of patients. Moreover, EHRs are expected to play an important role in connecting the scholarly information map [1] with nurses' information behavior that we previously proposed.

The purpose of this pilot study is to analyse problems of EHR when it is used as a clinical decision tool for nurses.

Methods

The case records in this study were acquired from the "National Database of Medical Adverse Events" [2, 3] provided by the Japan Council for Quality Health Care on its website. Among the case records, this study selectively acquired data on adverse events related to nurses by using the keywords of "EHR" or "information system". The searched records were analyzed through a natural language processing using the software called Text Mining Studio version 5.1 (Mathematical Systems Inc., Tokyo, Japan).

Results

The number of the full text case records retrieved from the initial search was 138. Five cases that had no direct relation to EHR were removed.

The results of the text mining for abstraction about EHR are displayed in Table 1. The actions of individual nurses that were considered to cause an adverse event were mainly identified as: failure to doublecheck, poor collaboration among health professionals, poor judgement on events in clinical settings, failure to observe patients' conditions.

Table 1– Examples of Frequent Words in the Records

Item	Words Related to EHR
Contents of Adverse Events	check, input, stop the prescription, open, injection
Background Factor	check, input, allergy, order, patients' profile, record, identification, system
Preventive Measures	input, check, identification, bulletin board, information

Conclusion

We analyzed medical adverse events data related to EHR in nursing services. More data that detail medical adverse events related to nursing services must be collected to create a meaningful linkage between EHR data and scholarly information. When that is realized, EHRs with comprehensive data will be a powerful clinical decision tool for nurses.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP26463280. We would like to thank the Japan Council for Quality Health Care for permitting us to use its database.

References

- [1] Mika Tomita, Mariko Iwasawa. A model for nurses seeking information using a scholarly information map. *Studies in Health Technology and Informatics*. 192 (2013), 1192
- [2] Japan Council for Quality Health Care. National Database of Medical Adverse Events. <http://www.med-safe.jp/contents/english/index.html> (2017-4-12 accessed)
- [3] Japan Council for Quality Health Care. National Database of Medical Adverse Events. <http://www.med-safe.jp/mpsearch/SearchReport.action> (2017-4-12 accessed)

Address for correspondence

Mika Tomita, Department of Nursing, School of Health Sciences, Ibaraki Prefectural University of Health Sciences, 4669-2, Ami, Amimachi, Inashiki-county, Ibaraki, Japan ; e-mail : tomitam@ipu.ac.jp

Using SNOMED Distance to Measure Semantic Similarity of Clinical Trials

Duo (Helen) Wei^a, Gang Fu^b

^a Computer Science and Information Systems, Stockton University, Galloway, NJ, USA,

^b Google Research, Google Inc., New York, NY, USA

Abstract

The objective of this research is to propose to develop an innovative distance function, called SNOMED distance, which captures the nature of the semantic distance within the topological structure of SNOMED, to identify semantic similarity between clinical trials.

Keywords:

Systematized Nomenclature of Medicine (SNOMED), Semantics, Clinical Trials

Introduction

One of the most widely used methods to group similar trials is called “clustering.” Significant research efforts have been devoted to produce clusters with low intra-cluster distance (high intra-cluster similarity) and high inter-cluster distance (low inter-cluster similarity). There are quite a few structural similarity measures between concepts used in natural language processing, and the Jaccard similarity coefficient [1] is considered the best one for calculating pair-wise structural similarity as it can assess both similarity and diversity. However, the semantic similarities among clinical trials are not adequately studied. In this research, we propose to develop an innovative distance function, called SNOMED distance, which captures the nature of the semantic distance within the topological structure of SNOMED, to identify semantic similarity between clinical trials. We expect that the application of SNOMED distance would lead to more reasonable trial clusterings.

Methods

The proposed SNOMED distance that captures semantic distance between concepts. Here “semantic distance” refers to

the relative closeness between two concepts of interest from a terminology or concept-oriented view. The SNOMED distance we proposed illustrated in Figure 1, where three cases (a) (b) (c) are considered. In particular, case (b) is a special case of (a).

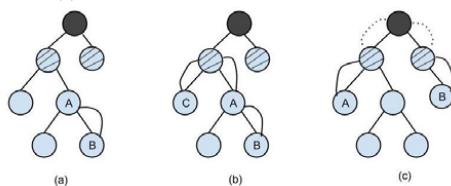


Figure 1. Illustration of SNOMED Distance in SNOMED terminology topological structure

We use filled gray circle to represent the root of SNOMED terminology called “SNOMED CT Concept”. The shaded circles represent the root of a hierarchy. Un-shaded light blue

circles represent SNOMED concepts. In the case of multiple paths connecting two concepts, only the shortest path will be considered. As shown in Figure 1(a), two nodes A and B are within the same hierarchy, and the SNOMED distance is defined as the shortest path between A and B. Therefore, it is one. Figure 1(b) is a special case of (a) when we calculate the SNOMED distance between B and C, because the shortest path must contain the lowest common ancestor of B and C. In this case is the hierarchy root (the shaded node), therefore, the SNOMED distance between B and C is three. Figure 1(c) represents the case that two nodes A and B reside in different hierarchies, and their SNOMED distance is close to infinity

Given two clinical trials, each of which is represented by a set of SNOMED concepts, such as Clinical Trial 1 (CT_1): $S = \{s_1, s_2, \dots, s_m\}$ and Clinical Trial 2 (CT_2): $T = \{t_1, t_2, \dots, t_n\}$. We calculate the distance between these two clinical trials using SNOMED distance (sct_dist) as follows.

$$Distance(S, T) = Median(\{d \mid d = sct_dist(s_i, t_j), s_i \in S, t_j \in T\})$$

Results

Let us take the following two clinical trials as examples. Assuming CT_1 is represented by {Electronic dental anesthesia, Pregnant woman, Rheumatic heart disease}, and CT_2 is represented by {Long menstrual cycle, Heart disease, Woman}. The SNOMED distance between each pair of SNOMED concepts appearing in two different clinical trials are shown in Table 1. Due to the space limitation, we are not able to show the SNOMED distance for all pairs of concepts. It is calculated that the SNOMED distance $sct_dist(CT_1, CT_2)$ is 12, $sct_dist(CT_2, CT_3)$ is 10012, and $sct_dist(CT_1, CT_3)$ is 10012.

Table1 Example SNOMED distance between pairs of SNOMED concepts in CT_1 & CT_2

CT_1	CT_2	$sct_dist(CT_1, CT_2)$
Electronic dental anesthesia	Long menstrual cycle	12
Electronic dental anesthesia	Heart disease	8
Electronic dental anesthesia	Woman	9
Pregnant woman	Long menstrual cycle	10014
Pregnant woman	Heart disease	10011

Conclusion

One limitation of the research is that the comparison between the proposed methods with existing distance function should be compared. We would like to show that SNOMED distance function can surpass existing distance function for the clustering in terms of accuracy and efficiency.

[1] Tanimoto, T. "An Elementary Mathematical theory of Classification and Prediction". Internal IBM Technical Report. 1957

Address for correspondence

duobwin@gmail.com

Terminology Adoption in Family Medicine Service – Start of Journey

Austen Wong^a, Maggie Lau^a, Vicky Fung^a, Dorothy To^b

^a Health Informatics Section, Hospital Authority, Hong Kong, China

^b Tuen Mun Hospital, Hospital Authority, Hong Kong, China

Abstract

A proposal on terminology adoption for the Problem and Encounter Reason capturing in the Hong Kong Hospital Authority Family Medicine Module was suggested. Given the complexity of the project, mapping had been conducted as preparatory work in the early phase of development.

Keywords:

Primary care, Standardisation, Mapping

Introduction

With the popularity of advanced information technology and informatics development in recent decades, clinical data documentation should be tailored to suit clinicians' workflow so as to offer them conveniences for facilitating their work. The Family Medicine Module (FM Module) in Hong Kong Hospital Authority (HA) was developed more than 20 years ago. A plan on reforming the Problems and Encounter Reasons capturing, with terminology adoption in particular, was suggested.

Methods

Implemented in 1995, the FM Module is a clinical system currently being used in more than 80 General Out-patient Clinics, Family Medicine Specialist Clinics and Staff Clinics within the territory under HA's management. Problem Oriented Medical Record (POMR) was adopted, which patients' individuality was emphasised by the list of problems specific to the person. Encounter Reasons for each episode of care would also be documented. Free text entries were being captured as Problems while International Classification of Primary Care-2 (ICPC-2) was adopted for Encounter Reason input.

In relation to the advancement in terminology development in HA for the past 10 years, review was conducted on the data capturing approach in FM Module. It was suggested to transform the Problem and Encounter Reason capturing to a complete adoption of Hong Kong Clinical Terminology Table (HKCTT). HKCTT embraces international reference terminologies and has been used for diagnosis and procedure reporting by other clinical modules within HA since 1995. It is sensible to align the terminology used in FM with the main stream in order to achieve interoperability within the organisation.

In the preparatory stage, a mapping of ICPC-2 terms to HKCTT had been conducted.

Results

After studying on the data distribution of Encounter Reason captured in the past, all 686 items in ICPC-2 Component 1 (Symptoms and complaints) and Component 7 (Diseases and diagnoses) were added into the HKCTT. As HKCTT is concept-based, ICPC-2 items that refers to a single clinical meaning were being handled differently with those carrying multiple meanings.

For ICPC-2 code which referred to a single clinical meaning, an attempt of locating a HKCTT concept with equivalent meaning was made. New HKCTT concepts with status indicator being "Active" were added for instances where the meaning being represented by the ICPC-2 item was not readily present, which denoted that the concept had been made available for future data capturing.

For ICPC-2 code that corresponded to more than one meaning, a new HKCTT concept with status indicator being "Inactive" was created. Further data capturing with that newly created concept would not be plausible. Nevertheless, the concept creation was necessary for preserving the processibility of historical data documented with ICPC-2. HKCTT concepts for individual split part of those "combined concepts" were created with concept status "Active" when an equivalent concept was not readily present in the table.

Table 1 shows the number of HKCTT concept mapped from a single ICPC2 item. 510/686 (~74%) of ICPC2 items were mapped to a single HKCTT concept.

Table 1 – ICPC2 to HKCTT mapping Summary

1 ICPC2 item map to	Total
1 HKCTT concept	510
2 HKCTT concepts	34
3 HKCTT concepts	120
4 HKCTT concepts	18
5 HKCTT concepts	4
Total no. of ICPC2 item	686

Conclusion

Patients care always comes first. Clinical documentation should be designed as seamless as possible. Yet, balance between convenience and data quality should be struck. Structured data should be adopted in view of the reusability that demonstrated. The enriched HKCTT would serve as a good fundamental for more work to come on the future development of the FM Module.

Application of Information Technology in the Outpatient Service Optimization

Zhang Xiaoying

School of Information, Nanjing University of Traditional Chinese Medicine, Nanjing, Jiangsu, China

Abstract

In a hierarchical diagnosis and treatment policy local tertiary hospitals assume the majority of clinic service, improving patient medical experience and enhancing service quality. Information technologies such as comprehensive self-services, and palm medical APPs can help solve these problems.

Keywords:

Quality of Health Care; Service optimization; Delivery of Health Care; Telemedicine.

Introduction

The Outpatient clinic is the first hospital window directly connected to the hospital's image [1]. The National Health and Family Planning Commission of the People's Republic of China has explicit requirements for tertiary hospitals: enhance working efficiency of outpatient clinic, optimize outpatient medical service procedures, shorten medical-therapy-seeking time, and increase visit satisfaction. Information technology is an efficient means to strengthen the hospital's scientific management and enhance the medical quality. Optimizing the outpatient clinic treatment procedure can help solve the "three longs and one short". Here, we monitor service quality promotion in the outpatient clinic by comparing the original flowchart to a new flowchart.

Methods

Service Optimization of Outpatient Clinic

Implementing the All-Purpose Card at the Outpatient Clinic. The all-purpose card integrates patients' basic information. It enables patients to pay a security deposit in advance, complete registration, make payments, receive diagnoses/treatments, get medications, and check inspection reports.

Time-Segmented Appointment-Making and Registration. Average treatment time is computed by the information system, which can be used to reduce registration and wait time.

Outpatient Clinic Payment/Appointment Implementation. Doctors can directly accept payments, care for patients, and schedule appointments reducing patient time in queue.

Carrying out Multiple Self-Services

To optimize the service and reduce the occurrences of "three longs and one short" multiple self-services are rendered.

Adopting Multiple Registration Forms. Appointment or registration by cellphone, website, official Wechat, on-site self-feeder, and manpower provide can reduce operational costs.

Self-Help Medicine Getting. Self-help medicine is used to help patients make payments and receive medications, reducing the time to fill prescriptions.

Self-Service: Result of Laboratory Chemical Examination. Self-service reporting machines, set outside of clinical

laboratories provide inspection and impact reports reducing reporting errors and patient wait time when receiving results.

Self-Service Evaluation System. Patients can readily evaluate the hospital service improving hospital service quality.

Electronization of Outpatient Clinic/ Information Sharing

Electronic prescription, medical record, and application forms can improve fee collection transparency and patient assurance.

Supporting Multiple Fee Charging Methods

Make payments with all-purpose card or bank cards or cellphone APP.

Mobile Health

Palm medical therapy has multiple integrated functions: registration/fee collection, online result access, payments, expert introduction, online consulting, satisfaction evaluation, and hospital maps.

Results

In the original flowchart of the outpatient clinic service process patients spent too much time in registration, waiting for treatment, making payments, filling prescriptions and waiting for results. The new system reduced treatment time and improved patient satisfaction and hospital service quality.

Conclusion

The optimization of outpatient clinic service by information technology shortens medical-therapy-receiving time, and enhances work efficiency, and quality of hospital outpatient clinic service solving the "three longs and one short" phenomenon rendering a superior medical-therapy-seeking experience.

References

- [1] Z. Qingying, Z. Cheng, Effect of city level hospital implementing diagnosis and treatment service by making appointment in advance, *J Nurs Admin* 14 (2014):375-376.
- [2] R. Guohong, Design and application of hospital all-purpose card advance payment account trusteeship, *China Medical Equipment* (2015):146-148.
- [3] W. Chong, S. Bo, G. Hauilei. Brief Analysis of the Application and Improvement of the Outpatient Clinic Information System, *Modern Hospital* 13 (2013):134-136.

Address for correspondence

Zhang Xiaoying

School of Information, Nanjing University of Traditional Chinese Medicine, China (210029)

Email: 18626003377@163.com; Phone:17721515776

An Open Metadata Schema for Clinical Pathway (openCP) in China

Wei Xu, Yanxin Zhu, Xia Wang

Department of Medical Informatics, Second Military Medical University, Shanghai, China

Abstract

China has issued and implemented standard clinical pathways (Chinese standard CPs) since 2009; however, they are still paper-based CPs. The aim of the study is to reorganize Chinese standard CPs based on related Chinese medical standards, by using archetype approach, and develop an Open platform for CP (openCP) in China.

Keywords:

Clinical pathway; archetype; data element.

Introduction

By the end of 2011, 331 standard CPs in 22 specialties have been developed in China and implemented in more than 3000 hospitals since a trial version of the Guiding Principle of CP was issued in October 2009[1].

However, Chinese standard CPs are still paper-based CPs, which hindered efficient development and use and was difficult to adapt to computer-based or even web-based environment[2].

The aim of the study is to reorganize Chinese standard CPs based on related Chinese medical standards, by using archetype approach, and develop an Open platform for CP (openCP) in China.

Methods

We chose 43 diseases covering 7 departments of internal medicine as samples and source materials including Respiratory, Gastroenterology, Neurology, Cardiology, Renal, Endocrinology, and Hematology, and we followed the 4 steps in reorganizing Chinese standard CPs and developing openCP: reorganize and develop the structure of Chinese standard CPs, add data groups, structure and standardize data elements, and create openCP.

Results

A four-level structured CP information model has been developed, basic data elements have been structured and standardized according to Chinese health related standards, creation rules for data elements have been set, and an Open platform for Clinical Pathway (openCP) has been built.

Structure of CP information model

CP information model in openCP has four hierarchical levels including Document, Section, Data Group, and Data Element. Document contains the top level structure, the CP, which consists of metadata about CP identifier, access, and status and top level sections for CP content.

Sections in openCP include Basic Information, Main Care Activities, Main Orders, and Main Nursing from original Chinese standard CPs.

Basic Information Section presents the state or situation of the patient, which includes demographic information, patient ID, diagnoses, procedures, implementation time, and etc.

Main Care Activities Section contains a series of care plans, observations, actions, and assessments. Most of data groups have corresponding ones in Chinese EHR Standard.

Both Main Orders Section and Main Nursing Section can reuse data groups and elements in Section of Main Care Activities.

Structure and standard of CP data elements

Data items in original Chinese standard CPs has not structured or standardized. If the data item is generic and has corresponding data element in Chinese health related standards, we adopted the data element directly. If there were no corresponding data elements, we created new names for the data items according to WS/T 303-2009 (Rules for data element standardization)[3]. As to the specific data items dedicated to CPs of specific diseases, we divided them into 2 categories and standardized them respectively. One can be transformed into common data elements plus specific data value. The other can not be transformed into common data elements, and we standardized them as specific data elements used for CPs of particular diseases.

Discussion and Conclusion

OpenCP has been developed and local CPs can be easily created or revised based on Chinese standard CPs. Adopting philosophy of openEHR archetype, we reorganized Chinese Standard CPs. The reorganization involved two steps: optimizing the structure of CP information model and standardizing the data elements. CP developers can create a CP by extracting data elements from these modules or add new data element according to the creation rules. It is an efficient way to develop a standardized, modularized and computerized CP.

Acknowledgements

The study is supported by National Social Science Foundation of China (13BTQ012).

References

- [1] The national assessment of clinical pathway pilot work. National Health and Family Planning Commission; Available from: <http://www.nhfpc.gov.cn/tigs/s9661/201206/c3d1c5948eb947ad944e0e96b1d1734f.shtml> (available:2016-10-11)
- [2] W. Xu, Z. Wang, Y. Zhua and et al. Analysis and Evaluation of Clinical Pathways in China. *Stud Health Technol Inform* (2013), 192:919.
- [3] Former Ministry of Health China.WS/T303-2009 Rules for data element standardization of health information [S]. 2009-01-22

Address for correspondence

Wei Xu Email: zhuizhui@smmu.edu.cn; Address: Dept. of Medical Informatics, Second Military Medical University, Shanghai, China.

Integration of Japanese Medical Device Adverse Event Terminologies

Ayako Yagahara^{a,b}, Takumi Tanikawa^{a,b}, Katsuhiko Ogasawara^b, Hideto Yokoi^c

^a Faculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan,

^b Faculty of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan,

^c Department of Medical Informatics, Kagawa University Hospital, Kida-gun, Kagawa, Japan

Abstract

The purpose of this study is to integrate Japanese medical device adverse event terminologies for evaluating terminological consistency. We represented hierarchy and relations among terms using Resource Description Framework (RDF). There were 3521 classes and 14650 properties. As a result of evaluating the consistency of the description in SPARQL, it was evident that the same notations existed within different terminologies (category terms and terms) and some terms had plural definitions.

Keywords:

Terminology; Equipment and Supplies; Medical Errors

Introduction

The Ministry of Health, Labour and Welfare announced the official Japanese medical devices adverse event terminologies (JMDAET) in March 2015 [1]. These are 90 terminologies for each medical device developed by 13 groups in the industry, who are members of The Japan Federation of Medical Devices Associations. The current terminology categories that have been announced are medical device problems, patient problems, and components. The 13 industry groups independently constructed each terminology by bottom-up method, conversely international terminologies were developed by top-down method. JMDAET was constructed to facilitate communication between users at all levels. The purpose of this study is to integrate these 90 terminologies to evaluate their terminological consistency.

Methods

The subject of this research included the categories of medical device problems and patient problems. The 90 terminologies recorded on separate Excel sheets were integrated onto one CSV file. The field names in CSV were as follows: category of terminology, terminology ID (which represented the industry and medical devices), category terms, preferred terms, definition, synonym, FDA code, and FDA source PT. We described the relations to fulfill the variables in the following fields using Resource Description Framework (RDF) (Figure 1). We used Google Refine 2.5 and Protégé 5.1.0 to describe and represent the relations. Finally, we evaluated the consistency of description in SPARQL, an RDF query language.

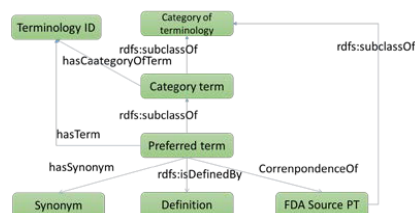


Figure 1 – Architecture of the relationships in Japanese medical device adverse event terminologies

Results

There were 3521 classes and 14650 properties. The maximum number of hierarchies, when the hierarchy between the category terms and preferred terms was represented, was 4. This meant that the same notation existed in category terms and preferred terms under different terminologies. In addition, we proved that the same notations had a maximum of 7 definitions. There were 109 terms that were both a category term and a preferred term, and 88 that were both a preferred term and a synonym. We considered integrating the defined representations or dividing the terms to ensure consistency. We could easily identify inconsistencies by visually representing the integrated terminology using SPARQL. It may be a useful tool to discuss the revisions in terminology among industry groups and to validate the consistency of the used terms.

Conclusions

In this study, we integrated JMDAET and found some points of inconsistency. Based on the results, we will consider editing JMDAET for consistency. In future work, this method may be used for consistency evaluation of terminology mapping between international terminologies and JMDAET.

Acknowledgements

This research is supported by the Research on Regulatory Science of Pharmaceuticals and Medical Devices from Japan Agency for Medical Research and Development, AMED.

Reference

- [1] The Japan Federation of Medical Devices Associations. Utilization of Japanese medical devices adverse event terminologies. <http://www.jfmda.gr.jp/fuguai/> (accessed 2017.7.3)

Address for correspondence

Ayako Yagahara yagahara-a@hus.ac.jp, Tel: +81-11-681-2161
Address: 7-Jo 15-4-1 Maeda, Teine, Sapporo, Hokkaido, 006-8585, Japan.

Infrastructure for Big Data in the Intensive Care Unit

Javier Zelechower^a, José Astudillo^a, Francisco Traversaro^a, Francisco Redelico^b,
Daniel Luna^c, Fernan Quiros^c, Eduardo San Roman^a, Marcelo Risk^{b,c}

^a Service of Intensive Care Unit, Hospital Italiano de Buenos Aires, Argentina

^b Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

^c Department of Health Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

The Big Data paradigm can be applied in intensive care unit, in order to improve the treatment of the patients, with the aim of customized decisions. This poster is about the infrastructure necessary to build a Big Data system for the ICU. Together with the infrastructure, the conformation of a multidisciplinary team is essential to develop Big Data to use in critical care medicine.

Keywords:

Intensive Care Unit, Vital Signs

Introduction

The application of Big Data [1] to the Intensive Care Unit [2], involves of both infrastructure and team work, this poster is about the first subject. The current technology of the monitor devices in the ICU, allows the capture of the vital signs beside the bed of the patient, and this monitors are capable of being connected to the computer network of the hospital. The objective of this poster is to propose a system for the infrastructure to build a Big Data system for the intensive care unit.

Methods

There are a number of fundamental issues that were resolved in order to carry out this kind of project: connectivity of equipment, internal communication system, control center, and server database. The figure 1 shows the diagram of the infrastructure, where the monitors are connected to a router, then the information flows through the network of the hospital, in the other end are connected the database server of the vital signs and the server of the EHR.

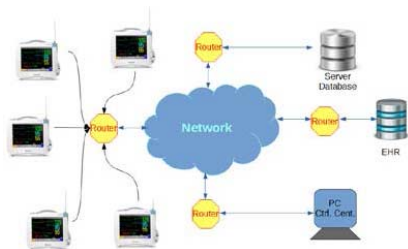


Figure 1– Infrastructure for Big Data in the ICU.

Among the components of the infrastructure the connectivity of the monitors is fundamental. In the absence of a standard communication protocol, each manufacturer has its own protocol, which means having to develop a specific interface for each of them. This part is one of the most complex to

solve, since the information provided for the manufacturer is scarce. For this reason, in this study a multi-parameter monitor was selected due to its highest quality information gathered.

In order to generate the information volume, the measurement is repeated in a previously selected sample of patients. The internal communication system is solved using the network of the hospital, owing to the capacities that it possesses with respect to the transmission of large amounts of information. Through the use of routers and with a planned subnetting, it is possible to identify each of the components in the network. There is a computer in charge of functioning as the Control Center. In this computer runs a processes to capture the data coming from the monitors. These processes analyze each captured data, perform its relevant decoding, and give them an appropriate format suitable for storage.

Due to the large volume of data that is generated by each patient, it is necessary to work with a dedicated database. For this reason there is a server located in a data center, in which it is stored, under a structure preferably designed, all the acquired information. Thus, allowing that this information can be later collected for subsequently analysis. This database also receives additional information from an electronic medical record, thus providing a large amount of static data. It must be taken into account that the personnel in charge of the monitors is instructed in order that the measurements of the parameters are not interrupted or are wrong, thus achieving the best possible quality of the acquired data. The database server of the vital signs runs a MongoDB software, this database was used because is non relational and therefore suitable for the kind of data generated by the monitors.

Discussion

The infrastructure propose in this paper has proven to be useful to store large amount of data from the monitors in the ICU. Then this data combined with the EHR, and analyzed by a multidisciplinary team, may help to a better assistance of the patients in the ICU.

References

- [1] Gandomi, A., and Haider, M., Beyond the hype: Big data concepts, methods, and analytics, *Int. J. Inf. Manage.* **35** (2015) 137–144.
- [2] Anthony Celi, L., Mark, R. G., Stone, D. J., and Montgomery, R. A., “Big Data” in the Intensive Care Unit. Closing the Data Loop, *Am. J. Respir. Crit. Care Med.* **187** (2013) 1157–1160.

Address for correspondence

Marcelo Risk, PhD, ScD. Hospital Italiano de Buenos Aires, Buenos Aires, Argentina. E-mail: marcelo.risk@hospitalitaliano.org.ar

Information Model Construction of EBM Resources

Yanxin Zhu, Wei Xu, Xia Wang

Department of Medical Informatics, Second Military Medical University, Shanghai, China

Abstract

To build evidence-based medicine (EBM) databases in China, we developed a semantic structure of EBM database information model by top-level, middle-level, and bottom-level structure. Top-level structure was composed of three modules including Literature Characteristics, Treatment Process and Evidence Levels. We further specified each top-level module based on international and national metadata and health related standards. Finally, we developed a complete information semantic model of EBM resources by an ontology tool. It can provide a reference for semantic construction of Chinese EBM database.

Keywords:

Evidence-Based Medicine, Metadata, Semantics

Introduction

It is often difficult for Chinese medical workers to obtain useful resources due to the lack of useful EBM databases. Our study is to establish a semantic structure of EBM information model, to better manage and organize the clinical literature which is a kind of important clinical evidence and to provide better EBM information service for medical personnel.

Methods

We used archetype method to construct complicated medical information into separate modules, whose elements were selected by literature guarantee principle. We built three combinable multi-level archetypes, and formed a complete information semantic model of EBM resources based on ontology.

Results

Top-level structure

EBM information needs of medical staff are multi-dimensional. In addition to the traditional but leading information needs known as literature characteristics, there's also needs for EBM evidence level and treatment process [1]. Therefore, we constructed the top-level structure into three separate modules: Literature Characteristics, Treatment Process, and Evidence Levels.

Middle-level structure

In the module of Literature Characteristics, we adopted *Dublin Core Metadata Element Set* [2] to organize and describe the middle-level semantic structure for better information sharing. Within the 15 data elements in Dublin Core, we chose and standardized 6 commonly used data elements such as title, description, source, language, subject, and creator.

In the module of Treatment Process, the data elements were mostly obtained from requirements research [1] and well-

known EBM databases abroad, such as disease, ICD-10, diagnosis, treatment, drug, symptoms, etiology, prognosis, patient information, etc. For future integration to EMR system in China, we also consulted *Date group and data element dictionary of CDA of electronic medical record* for reference.

In the module of Evidence Levels, we selected data elements such as trial design, intervention, outcomes, sample size, sequence generation, allocation concealment mechanism, blinding and diagram, based on randomized controlled trial (RCT) standard CONSORT 2010 [3].

Bottom-level structure

In bottom-level structure, we described each data element by Chinese standard WS363.1(*Health data element directory-Part1: General specification*) [4], which is the standard for medical information data element attribute description.

Conclusion

We developed a complete information semantic model of EBM resources based on international and national metadata and health related standards. It can provide a reference for semantic construction of Chinese EBM database.

Acknowledgements

The study is supported by National Social Science Foundation of China(13BTQ012).

References

- [1] Yanxin Zhu, Wei Xu. Investigation and analysis on information demands of medical staffs, *Journal of Medical Informatics(Chinese)* 36.3(2015), 62-66.
- [2] Dublincore.org [Internet]. USA: Dublin Core Application Profiles; c1995-2016[updated 2016 Oct 10; cited 2016 Oct 15]. Available from: <http://dublincore.org/documents/2009/05/18/profile-guidelines/>.
- [3] Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement, *JAMA* 276.8(1996):637-9.
- [4] Former Ministry of Health China.WS363.1 Health data element directory-Part1: General specification [S].2011-08-02.

Address for correspondence

Wei Xu (Email: zhuizhui@smmu.edu.cn)
Address: Dept. of Medical Informatics, Library, Second Military Medical University, Shanghai, China. Zip: 200433

Yanxin Zhu (Email: yanxinzhu@smmu.edu.cn)
Address: Dept. of Medical Informatics, Library, Second Military Medical University, Shanghai, China. Zip: 200433

This page intentionally left blank

V. Quality and Safety, and Patient Outcomes

This page intentionally left blank

A Hospital Nursing Adverse Events Reporting System Project: An Approach Based on the Systems Development Life Cycle

Yingjuan Cao^a, Marion Ball^b

^a Department of Nursing, Qilu Hospital of Shandong University, Jinan, Shandong, China

^b Johns Hopkins University, Senior Advisor IBM Research, Baltimore, Maryland, U.S.A

Abstract

Based on the System Development Life Cycle, a hospital based nursing adverse event reporting system was developed and implemented which integrated with the current Hospital Information System (HIS). Besides the positive outcomes in terms of timeliness and efficiency, this approach has brought an enormous change in how the nurses report, analyze and respond to the adverse events.

Keywords: Informatics; Nursing; Patient Safety

Introduction

An adverse events reporting system facilitating to respond to the events in a timely and efficient way is profound in enhancing patient safety and nursing care in all levels of health care institutes [1]. Many hospitals in developing countries like China are encountering a transitional phase in terms of building up a comprehensive and advanced hospital based medical informatics system. It is reported that the nursing adverse events reporting system is under developed which serves as one of the main barriers in promoting a patient safety culture among hospital staff in China [1]. To this end the Systems Development Life Cycle (SDLC) has been developed and implemented to ensure quality systems design and implementation [2]. This study was undertaken to develop and implement a hospital based nursing adverse events reporting system integrated with the current Hospital Information System (HIS).

Methods

This was a prospective study conducted in a tertiary teaching hospital in Jinan, P.R.China.

The SDLC was used to develop a nursing adverse events system in the study setting. The implementation process started in early 2013 and has followed the SDLC recommended plan. First, a team was assembled which made up of a few nursing unit managers and bed-side nursing staff as well as members of the hospital's quality and patient safety nursing subcommittee. Consequently, the functional requirements of the upcoming adverse events reporting system was assessed according to the hospital's local and national protocols and guidelines. This approach was followed by selecting a design approach, specifying the system's requirements.

Results

A nursing adverse events reporting platform was accomplished by working with the Information Technology (IT) company engineers and front-line nursing staff and managers, installing

and training, operating the system, and continuously evaluating and improving the system to achieve the defined goals and system requirements.

The average time saved to report an adverse event via the system, compared with the traditional filling in and hand-in way was more than 50%. An additional finding as a consequence of the systems implementation in 2014, has been a dramatic increase in the number of reported adverse events by nursing staff due to more accurate data now available. There are six main modules including:

- Patient demographic characteristics
- History of disease
- Risk assessment (e.g. falls, pressure ulcers, unplanned extubation, venous thromboembolism)
- Adverse events report
- Root cause analysis and action plan, and
- Statistical analysis data

Conclusions

The reporting system has brought an enormous change in how the nurses report, analyze and respond to the adverse events. Using this approach has made it possible for different levels of nursing managers, administrators, and bed-side nurses to detect those patients who are at high risks of adverse events, and to share and respond to adverse events in a timely and efficient manner.

References

- [1] Dai J, Wang Y, et al, *Investigation on the supervision pattern of benign interaction based on medical safety monitoring system*, Chinese Hospital Management, 34(2):28-29.
- [2] McBride S, Tietze M, *Nursing Informatics for the advanced practice nurse: Patient Safety, Quality, Outcomes, and Interprofessionalism*, Springer, 2015.

Address for correspondence

Yingjuan Cao, E: caoyingjuan@hotmail.com M: 18560082780

Data Quality of the Chinese National AIDS Information System: A Critical Review

Hong Chen^{a, b}, Ping Yu^a, David Hailey^a, Tingru Cui^a

^a Centre for IT-enabled Transformation, School of Computing and Information Technology,
Faculty of Engineering and Information Sciences, University of Wollongong, Wollongong, NSW, Australia,
^b Jiangxi Provincial Centre for Disease Control and Prevention, Nanchang, Jiangxi, P.R. China

Abstract

Thirty-nine electronic English and Chinese articles on data quality assessment of the Chinese AIDS information system were critically reviewed. Some performance assessment related indicators of data quality have improved since the system was launched in 2008. After a thematic analysis of the factors that may affect data quality, four domains were identified. They are data management, data collector, information system, and data collection environment. The findings are useful to guide data quality improvement effort.

Keywords:

Data quality; AIDS; information system.

Introduction

We conducted a critical literature review to explore data quality assessment in Chinese AIDS Comprehensive Response Information Management System (CRIMS): the status of and the influential factors of data quality.

Methods

English and Chinese electronic literature databases were searched, such as Scopus and CNKI with keywords “AIDS”, “data quality” and “China”. Thirty-nine articles were critically reviewed and thematically synthesized to conceptualize the factors that may affect data quality.

Results

Some performance assessment related indicators of data quality have improved since the CRIMS was launched in 2008. By 2013, the case follow-up rate and the case epidemiological survey rate were greater than 98.0%. The rate of data consistency in laboratory testing of CD4⁺ T cell counting was maintained above 90.0% except for 2010 [1]. However, non-compulsory items remained incomplete, e.g., identity number, workplace, and contact phone number [1, 3].

The factors associated with CRIMS's data quality were identified and grouped into 14 categories under four domains: data management, data collector, information systems, and data collection environment (see Figure 1).

Data management is of particular concern when the annual national data-driven performance assessment is used to compare data quality at all levels of the CRIMS [1]. A further concern is the data collection environment in which health service clients sometimes could inhibit data quality in data collection process. Inadequate communication with the clients by data collectors is reported. Automatic data entry checking function is welcomed in the field but may increase the uncertainty of data elements [2].

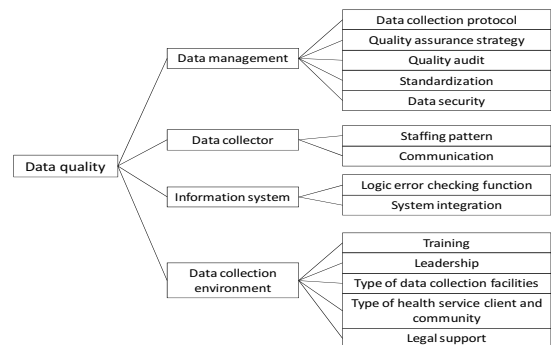


Figure 1 - Factors that may affect data quality in the CRIMS

Conclusions

High-level data quality for performance assessment related indicators was reported in the CRIMS, although the problem of the incompleteness of non-compulsory data remained by 2013. The factor affecting data quality can be grouped into four domains: data management, data collector, information system, and data collection environment. Further research needs to investigate how the four domains affect the quality of the data collection process.

Acknowledgements

This research has been conducted with the support of an Australian Government Research Training Program Scholarship.

References

- [1] J. Han, H. Tang, J. Xu, and Y. Mao, Analysis of national data quality assessment for follow up of HIV/AIDS cases between 2009 and 2013, *Chin J Prev Med* **48** (2014), 1104-1106.
- [2] J. Liu, Z. Jiang, W. Chen, G. Shan, Y. Fan, and H. Zhang, "Case not found" of HIV/AIDS cases analysis in Liuzhou City, 1996-2010, *Chin J Dis Control Prev* **16** (2012), 307-310.
- [3] Q. Qin, L. Wang, Z. Ding, L. Wang, F. Hei, P. Li, and F. Chen, Evaluation of AIDS network direct reporting quality and related indicators in China, 2010, *Disease Surveillance* **27** (2012), 1007-1012.

Address for correspondence

Ping Yu
Associate Professor, PhD
School of Computing and Information Technology,
University of Wollongong, Wollongong, NSW 2511, Australia,
Email: ping@uow.edu.au

Evaluation of Clinical Nursing Information System in Taiwan Regional Hospital

Wan-Jie Chou^{a,b}, Pei-Yu Tsai^{a,c}, Shu-Ying Lin^d, I-Ching Hou^a

^a School of Nursing, National Yang-Ming University, Taipei, Taiwan, China

^b Department of Nursing, Central Clinic Hospital, Taipei, Taiwan, China

^c Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan, China

^d Department of Nursing, Taoyuan General Hospital, Ministry of Health and Welfare, Taoyuan, Taiwan, China

Abstract

The purpose of this study was to evaluate the Clinical Nursing Information System (CNIS) in Taiwan regional hospital. In 2016, a total of 333 nurses responded to the Technology Acceptance Model-based questionnaire after 15 months of CNIS implementation. The results showed positive acceptance toward CNIS, especially among those nurses who were younger, those who worked as administrative managers or in non-critical care units, and had advanced computer skills.

Keywords:

Computer Systems, Nurses, Taiwan

Introduction

Clinical Nursing Information Systems (CNIS) improve the communication between health care team to enhance the healthcare quality and work efficiency. In June 2015, an 862-bed regional hospital in Taiwan implemented CNIS, including a total of 19 sub-systems such as pain assessment etc (see Figure 1). The purpose of this study was to investigate nurses' perception toward the CNIS and as a reference for the similar healthcare institutes in the future.



Figure 1. Snapshot of pain assessment

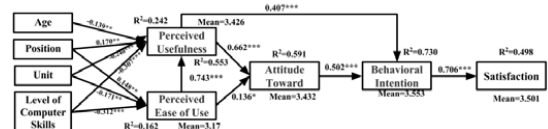
Methods

A cross-sectional study was conducted in 2016 after 15 months CNIS implementation. To collect data, a questionnaire consisted of 13 demographic variables, 52 Technology Acceptance Model (TAM) based items [1] with five-point Likert scales and 3 open questions was developed and validated (Cronbach α was 0.89). Half of the nurses were randomly sampled to fill out the questionnaire and there was a total of 333 valid respondents, the effective return rate was 99.4%.

Results

The respondents were mostly female (96.6%); 31.8 years old in average; 52.6% with bachelor degree; 67.6% work in non-critical care units, and 46.5% self-report the level proficiency in computer skills was intermediate. The average score of overall CNIS satisfaction was 3.50. Subjects were most satisfied with "CNIS helped me to save writing time" (3.72) and most unsatisfied with "screen switching, storage, delay-

prone and crash"(1.97). Nurses who were younger, as administrative managers, work in non-critical care units and had advanced computer skills were most satisfied with CNIS. Those 19 sub-systems were all reached at least 3.40 satisfaction score in average. Nurses' perceived usefulness, perceived ease of use, and attitude toward using all had a positive influence on the willingness of use and satisfaction (see Figure 2). The frequently narrative descriptions from open questions showed that CNIS reduced the writing time, increased efficiency and convenience, realized the patient dynamics and situations quickly, and increased the interaction between care team.



Conclusions

The results showed that nurses had positive acceptance toward CNIS. However, the stability of the system should be improved first. For those unsatisfied staff (e.g., critical care units), it will be recommended that more advanced information (e.g., qualitative interview) should be collected and analyzed in the future. Improving the problems occurred frequently could also help to satisfy users. Consequently, the implementation of CNIS was really beneficial to nurses' work efficiency and quality.

Acknowledgements

The research team thank Taoyuan General Hospital, Ministry of Health and Welfare for funding this study.

References

- [1] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, *Management Information Systems quarterly* 13 (1989), 319-340.

Address for correspondence

Corresponding Author: I-Ching Hou
Address: Room 407, School of Nursing, No.155, Sec. 2, Linong St., Beitou Dist., Taipei City 112, Taiwan
E-mail Address: evita@ym.edu.tw
Phone: 886-2-28267315

Personalizing Longitudinal Care Coordination for Patients with Chronic Kidney Disease

Theresa A. Cullen, MD, MS^a, Suranga N. Kasthurirathne, BEng^b,
Jenna M. Norton, MPH^c, Andrew S. Narva, MD^c

^a Regenstrief Institute, Indianapolis, IN, USA; ^b School of Informatics and Computing, Indiana University, IN, USA;

^c National Institutes of Health, Bethesda, MD, USA

Abstract

Chronic care coordination efforts often focus on the needs of the healthcare team and not on the individual needs of each patient. However, developing a personalized care plan for patients with Chronic Kidney Disease (CKD) requires individual patient engagement with the health care team. We describe the development of a CKD e-care plan that focuses on patient specific needs and life goals, and can be personalized according to provider needs.

Keywords:

Precision medicine; Renal Insufficiency, Chronic

Introduction

Chronic kidney disease (CKD) is a significant medical challenge that affects between 8-16% of the world's population [1]. CKD patients are often faced with multiple medical problems, the management of which may be impeded by poor interoperability of digital information and insufficient communication amongst providers and with patients. CKD patients are prescribed many medications, require frequent procedures, are at high risk for acute events, and experience frequent transitions across multiple care settings [2]. Due to its complexity and severity, CKD patients may hold different life goals that should significantly influence decision-making points that occur throughout their illness. In each case, it is essential to tailor treatment and care plans to meet patient specific life goals. The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), one of the National Institutes of Health of the USA, sought to personalize care for CKD patients via an e-care plan tailored to capture patient specific needs that could be transferred across multiple care sites.

Methods

The NIDDK's Kidney Disease Education Program convened an expert working group comprising of patients, nephrologists, and primary care physicians to facilitate the development of an e-care plan to facilitate a longitudinal transfer of key patient data across providers and health care settings using HL7 Consolidated Clinical Document Architecture (C-CDA) [3]. It also conducted a series of patient phone interviews to inform this process via the identification of significant decision points and CKD care goals of value to patients.

Results

We identified 55 indicators for inclusion in the CKD e-care plan. These indicators range from measurements such as serum creatinine levels to other predictive clinical factors, renal replacement therapy decisions, evaluation for mental

health status (including depression), functional status, activities of daily living and achievement of personal health goals. For each indicator, we assigned numeric and categorical priority scores based on the working group's recommendations. We also identified appropriate clinical terminology spread across LOINC, SNOMED, ICD-10 (diseases) and CPT codes (procedures). We found that many data elements essential to optimal CKD care lack appropriate clinical terms for information sharing across settings. This included specific patient focused terms such as choice for renal replacement therapy, patient education, patient goals and much more. Standard codes were used. Where standard LOINC codes were unavailable, we contacted LOINC to create new codes. A draft data set was published online [4] and welcomes feedback. Our next step is to pilot and test the data set at multiple settings.

Conclusions

A common data set is essential for developing an e-care plan. However, the plan must also have the flexibility to allow multiple views of the data, including a personalized patient perspective as well as functionality for individual providers to create personalized templates or dashboards that present the common set of data elements based on the needs of their clinical disciplines. The e-care plan must also find an appropriate balance between brevity and comprehensiveness. It should also be person, and not solely disease-centric. A disease-based focus could yield numerous, uncoordinated plans for people with multiple chronic conditions. Many data elements essential to optimal CKD care lack standards and codes for sharing information uniformly across settings. Additionally, global issues with interoperability, inadequate clinical terminology, inconsistent health information exchanges, and medication and other clinical reconciliation challenges pose problems for implementing the care plan.

References

- [1] V. Jha, G. Garcia-Garcia, K. Iseki, Z. Li, S. Naicker, B. Plattner, R. Saran, A.Y.-M. Wang, and C.-W. Yang, Chronic kidney disease: global dimension and perspectives, *The Lancet* 382 (2013), 260-272.
- [2] L.S. Chawla and P.L. Kimmel, Acute kidney injury and chronic kidney disease: an integrated clinical syndrome, *Kidney International* 82 (2012), 516-524.
- [3] HealthIT.gov, Consolidated CDA Overview, in, 2016.
- [4] National Institute of Diabetes and Digestive and Kidney Diseases, Development of an Electronic CKD Care Plan, in, 2016.

Address for correspondence

Suranga N. Kasthurirathne, BEng, School of Informatics and Computing (SOIC), Indiana University, Indianapolis, IN, USA: snkasthu@iupui.edu

A Guide for Constructing Bayesian Network Graphs of Cancer Treatment Decisions

Mario A. Cypko^a, Matthaeus Stoeck^b, Steffen Oeltze-Jafra^a, Andreas Dietz^b, Heinz U. Lemke^c

^aInnovation Center Computer Assisted Surgery (ICCAS), Medical Faculty, University of Leipzig, Leipzig, Germany

^bDepartment of Head Medicine and Oral Health, Head and Neck Surgery, University of Leipzig, Leipzig, Germany

^cInternational Foundation for Computer Assisted Radiology and Surgery (IFCARS), Kuessaberg, Germany

Abstract

In complex cancer cases, Bayesian networks can support clinical experts in finding the best patient-specific therapeutic decisions. However, the development of decision networks requires teamwork of at least one domain expert and one knowledge engineer making the process expensive, time-consuming, and prone to misunderstandings. We present a novel method for guided modeling. This method enables domain experts to model collaboratively without the need of knowledge engineers, increasing both the development speed and model quality.

Keywords:

Decision Support Systems, Clinical; Bayes Theorem; Expert Systems

Introduction

Bayesian networks (BN) has a great potential in patient-specific therapeutic decision making [1]. However, networks require knowledge from domain experts but they are usually neither BN experts nor able to use the often complex BN-specific modeling tools. Furthermore, BNs allow for very flexible knowledge representations; modeling attempts by two experts are likely to create different models despite a shared understanding of physiological mechanisms and causalities. Therefore, domain experts require teamwork with at least one knowledge engineer making the process expensive, time-consuming, and prone to misunderstandings. A computer system can supersede the knowledge engineer and enable an user-unspecific modeling.

Methods

A BN represents the decision model by a directed, acyclic graph (DAG) with conditional probabilities [1]. A DAG consists of nodes representing random variables and directed edges between nodes representing their direct causal dependency. An exemplary network for the treatment of laryngeal cancer is represented in Figure 1. The therapy decision network includes all relevant characteristics of the target disease, therapy options, and therapy effects. Based on a network and a set of observations (patient information), an inference algorithm computes the likelihood of occurrence for the network's remaining unobserved characteristics. The larynx model resulted from three-year teamwork of ENT clinician and knowledge engineer (both authors of this poster) [2]. The domain expert externalized medical knowledge and experiences, while the knowledge engineer translated it into a network structure.

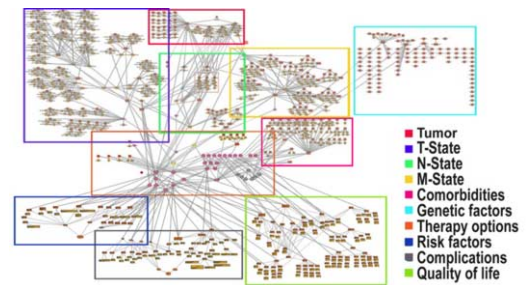


Figure 1 – Bayesian network of the therapy decision for laryngeal cancer, with over 900 variables and over 1200 dependencies [2].

Results

From the long term teamwork as well as weekly expert meetings, we elaborate a guided modeling approach tailored to clinical understanding. The general idea was, to transform complex and flexible BN modeling tasks into a natural language questionnaire from with a predefined modeling process. Our approach comprises 1) the questionnaire form facilitating a structured input of clinical knowledge, i.e. variables and their causalities 2) a nomenclature of clinical terms, supporting model exchange, comparison and search as well as distributed modeling and 3) a graph representation of the knowledge, which shows the modeling status and is automatically built as a background task.

Conclusion

With an appropriate web-based modeling framework, collaborative modeling between domain experts will be possible, would expedite the model development of other diagnosis and therapies, and allow for discussing domain-specific decision.

References

- [1] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, 1988.
- [2] M.A. Cypko, M. Stoeck, M. Kozniowski, M.J. Druzdzal, A. Dietz, H.U. Lemke, Validation workflow for a clinical Bayesian network model in multidisciplinary decision making in head and neck oncology treatment. *Int J Comput Assist Radiol Surg*, 2017.

Address for correspondence

Mario A. Cypko; email: mario.cypko@medizin.uni-leipzig.de

Modeling a System for Generating Structured Reports

Lorena Dutra da Costa, Marcos Cordeiro d'Ornellas

Department of Applied Computing, Federal University of Santa Maria, Santa Maria, Rio Grande do Sul, Brasil

Abstract

The purpose of this research is to make the medical report generation process more practical, fast and reliable, both for the health professional and for the patient. We created an ontology and modeling of a structured report (SR) Standard DICOM SR.

Keywords:

Reports, Radiology, Informatics

Introduction

The work of health professionals has been changing in recent years. The information previously stored on paper first consisted of digital files, with this, the need arose to normalize and standardize clinical information through reliable and efficient systems. Most healthcare institutions do not use SR, as a consequence, the management and retrieval of this information becomes a complex challenge that demands a lot of service time. This makes it difficult to extract relevant information and especially the comparison between reports, and the variability between structures used can easily lead to confusion and lead to medical errors. SR reduces the possibility of errors that could occur if performed with free text. The data output of these systems will be in a specific structure, since medical assistants use common terminology and pathways to generate reports and interpretation is performed in a guided way using that model used. The time for producing a report using defined templates is reduced, and the productivity of the radiologist increased intensified, thereby improving departmental [1].

Method

Create an ontology for structuring the content contained in radiological reports of computed tomography of the skull, and create models / reports templates that can be used in the system to be modeled. A bibliographical research was conducted on structured reports that sought the state of the art and the systems being developed as SR editors.

Result

The system to be modeled allows the creation of structured medical reports from models. The standard followed is the DICOM SR standard, an international standard used for storage and sharing of medical reports. Patient information is encoded in the form of objects that can contain text, DICOM images, audio files, and other information. Table 1 shows some studies in which systems have been developed for this purpose.

Table 1 – State of the art - Structure Report

Name	Objective	Ref.	Note
Institute of Telematics	View, edition and storage.	[2]	Not fully structured
Perceptive Express Reporter	Transforms reports into DICOM SR.	[3]	Paid
T-Rex	Creates and edits SR templates.	[4]	Open source

Conclusion

In the system modeled here each part of the final report created constitutes a subtree of the structure, the system constructs the final structure by automatically mounting the sub-trees according to preconfigured rules. The possibility of errors is diminished and therefore customer service is also improved. With the bibliographic research already done, it was possible to notice that many studies are being done about SR [5]. In most cases, comparisons are made between the use of SR and free text, the results show that the readability and linguistic quality of the reports are improved, leading to greater satisfaction among the professionals involved, but do not extend to until the development of system for creating SR.

References

- [1] J. E. Stewart. Radiologists Prefer Structured Reporting? 2015 Supplied in: <https://www.linkedin.com/pulse/radiologists-prefer-structured-reporting-john-e-stewart-md-phd>. Access in: 12/05/2016.
- [2] V. Lutz; M. Christoph. A Multimedia-Editor for making Findings in Radiology. Proceedings of the Thirteenth IEEE Symposium on Computer Based Medical Systems. Houston, Texas. 2002.
- [3] RASNA Imaging. Perceptive Reporter. Supplied in: <http://www.rasnaimaging.com>. Access in: 13/01/2017.
- [4] Karos Health . Supplied in: <https://www.karoshealth.com> Access in: 05/12/2017.
- [5] J. D. Segrelles et al. Increasing the Efficiency on Producing Radiology Reports for Breast Cancer Diagnosis by Means of Structured Reports. Methods of Information in Medicine. 2017.

Address for correspondence

Lorena Dutra da Costa, ldutra@inf.ufsm.br, (55) 5584286182.

Event Notification in Support of Population Health: The Promise and Challenges from a Randomized Controlled Trial

Brian E. Dixon^{a,b,c}, Kenneth S. Boockvar^{d,e}

^a Indianapolis VA Medical Center, U.S. Department of Veterans Affairs, Indianapolis, Indiana, USA

^b Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, Indiana, USA,

^c Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, Indiana, USA

^d James J. Peters VA Medical Center, U.S. Department of Veterans Affairs, Bronx, New York, USA

^e Icahn School of Medicine at Mount Sinai, New York, New York, USA

Abstract

Event notifications are real-time, electronic alerts that have the promise of improving population health by exchanging critical information to a patient's extended care team. In a trial of event notifications in U.S. Veterans Affairs facilities, we seek to understand the impact of notifications on health care utilization within 30 and 90-days. Lessons from the trial have implications beyond the evidence by informing strategies to develop and implement event notifications in other health systems.

Keywords:

Health Information Exchange; Organization and Administration; Veterans

Introduction

Event notification is the real-time, electronic, automatic alerting of providers to their patients' contact with other health care facilities [1]. Also referred to as an alert or subscription service, event notifications are messages triggered by specific patient activities, such as a visit being registered in a hospital's admission-discharge-transfer (ADT) system. Triggers often involve an inpatient admission, discharge, or emergency department visit. Event notification systems fall under the broader category of health information exchange (HIE).

Event notification holds promise for improving population health. Event notifications are a source of information about patients, thereby increasing providers' general awareness of patients' complex medical histories. Additionally, informing the provider by event notification creates opportunities for immediate intervention where appropriate. For example, care coordinators may be able to contact the emergency department prior to inpatient admission to provide insights that may prevent unnecessary utilization of resources. Likewise, if the patient is still at the emergency department or hospital, the ambulatory primary care (PC) provider could support the coordination of post-discharge transitions and services, potentially improving the quality of care for a patient. Event notification may also help identify patients for referral into care coordination programs.

The evidence base for the effectiveness of event notifications is just beginning to grow. Few studies have examined these systems in real-world settings.

Methods

We are conducting a cluster randomized trial in two geographically diverse medical centers within the U.S. Department of Veterans Affairs (VA) to examine the impact of event notifications for non-VA on quality of care following acute events. Older Veterans (≥ 65 years of age) are randomized

to one of two arms based on their PC provider: 1) event notifications to PC; or 2) event notifications plus a 30-day care coordination intervention designed to engage Veterans in managing their health. The trial seeks to understand the impact of even notifications and the intervention on 30-day and 90-day readmissions. We further seek to understand the acceptance of event notifications by patients and providers.

Results

The trial began in mid-2016 with an emphasis to date on patient recruitment and development of the event notification services. Event notifications are delivered to the VA using two different technical architectures. In the Bronx, New York, event notifications are delivered to a provider portal in which staff at the VA log into to view details about non-VA events. In Indianapolis, Indiana, notifications are delivered to VA systems using the Direct standard, and details of the non-VA events are entered as clinical notes in the VA electronic health record.

Conclusion

While event notification services hold much promise, to date there have been challenges to introduce them within the VA. Providers are wary about the utility of notifications as they perceive them as a new 'alert' that may interrupt workflow. Patients generally view them favorably, although many Veterans do not perceive the need for them as they do not believe they will consume non-VA care (an assumption not supported by available evidence). Lessons from this trial will be of use to other health systems that seek to leverage event notifications in support of population health.

References

- [1] T. Moore, J.S. Shapiro, L. Doles, N. Calman, E. Camhi, T. Check, A. Onyile, and G. Kuperman, Event detection: a clinical notification service on a health information exchange platform, *AMIA Annu Symp Proc* **2012** (2012), 635-642.

Address for correspondence

Brian Dixon, PhD, Health Research Scientist
1481 W. 10th St., 11H, Indianapolis, IN 46202 USA
+1 (317) 988-4419
brian.dixon@va.gov

Automatic Loading of Problems Using a Comorbidities Subset: One Step to Organize and Maintain the Patient's Problem List

Santiago Márquez Fossier, Alejandro Gaiera, Carlos Otero, Sonia Benitez, Daniel Luna, Fernán Quiroz

Departamento de Informática en Salud, Hospital Italiano de Buenos Aires

Abstract:

An accurate and updated problems' list is critical in a problem-oriented Electronic Health Record (EHR). The lack of organization and maintenance of the problems limits its value. Certain problems have a larger effect on the clinical evolution of the patient, these are known as Comorbidities. The aim of this paper is to evaluate the impact of the automatic loading of comorbidities in the organization and maintenance of inpatient problems' list using a comorbidities subset.

Keywords

Comorbidity; Medical Records, Problem-Oriented; Cost of illness.

Introduction

The problem of conceptualizing and organizing patient data in the electronic health record (EHR) around problems is a challenge [1]. An accurate and updated problem list is critical in a problem-oriented EHR. However, keeping a precise problems list can be a challenge; an inaccurate or incomplete problem list has deleterious effects on several patient care functions [2,3]. The comorbidities are problems that affect the clinical evolution of the patient and are associated with worsened health outcomes, more complex clinical management, and increased health care costs [4]. The objective of this study is to analyse the impact of automatically loading problems during the problems reconciliation in admission summaries for the organization and maintenance of the inpatient problem lists using a comorbidities subset.

Methods

Since 1998, the Hospital Italiano de Buenos Aires (HIBA) has run an in-house developed health information system. Its EHR is an integrated, problem-oriented, patient centered system recently certified by HIMSS as level 6+ [5]. All recorded problems use SNOMED CT. We use a previously developed set comorbidities [6] of 672 concepts in hospital admissions during problem reconciliation to allow automatic loading of problems. On 7/14/16, automatic loading of outpatient comorbidities was deployed on all admission summaries for reconciliation by the attending physician. A pre- and post- deployment analysis was performed based on the amount of structured problems.

We evaluated a period of 150 days pre- and post-implementation, calculated the mean of recorded problems, and analyzed the status modification (deletes/passed to error) in each hospitalization to evaluate the maintenance of the inpatient's problem list.

Results

A total of 39239 hospitalization were analysed. Patients had an

an average age of 47 years (21621 females) and the average length of stay was 2.74 days. There were 19577 pre-deployment hospitalizations (mean of 130 daily admissions), 20183 structured and coded problems (mean of 135 daily), with an average of 1.03 per hospitalization. There were 19662 post-deployment admissions (average of 131/day) and 92828 problems (mean of 619 records/day) with a mean of 4.72 daily. There was a 386% increase (316 pre/1219 post-development) in structure problems with professional status modification (passed to error/resolved), during hospitalization. The most registered pre-development problems were: High blood pressure (1042), Newborn (632), Labor (388), Scheduled Caesarean (302), Hypothyroidism (273), Diabetes Mellitus type 2 (232), Fiber (221). Post-deployment problems included: High blood pressure (6303), Dyslipidemia (3417), Tobacco Abuse (2247), Hypothyroidism (1637), Obesity (1293). The most frequent status modification pre-deployment were: High blood pressure (10), Rituximab (6), Bone marrow transplant (5), Hepatic biopsy (5), Fiber (4), New born to term (4), Stroke (4) Pneumonia (3), New born to term with adequate weight (3). Post-deployment: High blood pressure (83), Dyslipidemia (53), Tobacco abuse (38), Stroke (31), Ischemic stroke (30), Diabetes Mellitus (25), Valvulopathy (20), Cataract (20), Chronic renal failure (19).

Conclusion

The development of a comorbidities subset based on the WHO's Global Burden of Disease measure using disability-adjusted life years (DALY) lead to more meaningful/actionable problems being added to the problem and more frequent problem list reconciliation (i.e., removing inaccurate/outdated problems). We conclude that more studies are necessary to obtain a complete and adequate problem-oriented EHR.

References

- [1] Wright, *Clinical Problem Lists in the Electronic Health Record*, CRC Press, 2015.
- [2] D. Blumenthal, M. Tavenner, "The meaningful use" regulation for electronic health records, *N Engl J Med* **363** (2010): 501-504.
- [3] O. Shabestari, A. Roudsari, "Challenges in data quality assurance for electronic health records, *Stud Health Tech and Inform* **183** (2013):37-41.
- [4] Starfield, "Threads and yarns: weaving the tapestry of comorbidity, *Ann Fam Med* **4.2** (2006):101-103.
- [5] Lee, N. de Keizer, F. Lau, R. Cornet, "Literature review of SNOMED CT use, *J Am Med Inform Assoc* **21**, (2014):11-9.
- [6] S. Márquez Fossier (2016). Characterizing comorbidities, a step to organize the patient problems list, In: *Congresso Brasileiro de Informática em Saúde (SBIS)*, Goiana Brasil.

Address for correspondence

Santiago Márquez Fossier, Department of Health Informatics.

Juan D. Perón 4190, C1181ACH Buenos Aires, Argentina.

E-mail: santiago.marquez@hospitalitaliano.org.ar

Using Health Information Technology to Enhance Care Outcome Accountability Through Bundled Payments

Rima Artonian Gibbings^a, Nilmini Wickramasinghe^{a,b}

^a University of Illinois in Chicago, Chicago, USA,

^b Deakin University and Epworth HealthCare, Melbourne, Victoria, Australia

Abstract

Eventhough the US expenditure on health care is almost twice of the Organization for Economic Co-operation and Development (OECD) average, the care quality indicators in the US unfortunately are not reflective of such extra spending (Baird, 2016). To revise this cost and quality imbalance, new models have been recommended (and partially implemented) that will tie health care reimbursement with efficient outcomes. This study offers a technology enabled solution to calculating appropriate bundles.

Keywords:

Bundled payment

Introduction

The CMS's Innovation Center (IC) has initiated the Bundled Payments for Care Improvement (BPCI) initiative to measure the degree of success in improving quality of care and reducing cost with selected organizations [2]. In reviewing the several models suggested through the BPCI, the focus is to determine a retrospective bundled payment arrangement that will reconcile the actual rendered care services' cost with the CMS's targeted price for a given episode of care. In addition to CMS, twenty major health systems including payers, providers, patients, and purchasers have joined forces in forming an industry consortium called the Health Care Transformation Task Force.

Background

Health care expenditure is claiming 18 percent of the entire US GDP and is projected that by 2023 this share will experience a sharp rise of an additional 28% to consume the fifth of the nation's GDP [3]. Policymakers are scrambling to reduce "wasteful spending" by targeting administrative and clinical inefficiencies resulting in improved quality and accountability. Replacing fee-for-service with pay-for-performance (P4P) is aimed at rewarding health care facilities who provide higher quality of care with reduced cost. The main goals of P4P (value-based purchasing) is ensuring that patients receive adequate, timely, cost-effective and efficient care through improving quality of care, reducing complication episodes, and emphasizing on timely and preventive services. These steps are designed to reduce health care underuse and overuse. The Patient Protection and Affordable Care Act (ACA) introduced the "payment modifier" which allowed payment adjustment based on quality of care. In addition to introducing Accountable Care Organizations (ACO) who will stand to share risk together in providing treatment. CMS announced a landmark program known as outcome-based payment model for the Comprehensive Care for Joint

Replacement (CJR) that went into effect April 1, 2016 (bundled payment).

Method

The research method is based on a design science research methodology to develop a technology solution. The study is based on categorizing medical procedures based on predetermined CMS bundled payments. Further, it examines the role of developed technology solution to assist with promoting accountability in the VBP. The study focuses on enabling health care facilities to determine bundled payment encodings and reflects the codes in the payment process. Facilities will be capable in identifying the procedural charges of clinical processes according to CMS bundle settings. The outcome of facility payment categorization will improve strategic planning and provide standardization incentives. Health care facilities will be able to determine units that are continuously underperforming based on the new payment classification model.

Results

This is a research in progress study. The main idea in this study is to assist health care facilities in creating road maps prior to adopting bundled payments. The bundled payment model has only been adopted in portion of pilot hospitals so far. The accurate implementation of bundled payment will provide valuable knowledge to facilities that will identify critical changes to their service delivery methods based on the outcome. Currently the conceptual model and first parts of the technology solution have been developed.

References

- [1] Baird, C. (2016, March). Top Healthcare Stories for 2016: Pay-for-Performance. *Committee for Economic Development*. Retrieved from <https://www.ced.org/blog/entry/top-healthcare-stories-for-2016-pay-for-performance>
- [2] Centers for Medicare and Medicaid Services. (2016, November). *Bundled Payments for Care Improvement (BPCI) Initiative: General Information*. Retrieved from CMS.gov: <https://innovation.cms.gov/initiatives/bundled-payments/>
- [3] Chantrell, C. (2016, November). *Government Spending Chart*. Retrieved from USgovernmentsspending.com: http://www.usgovernmentsspending.com/spending_chart_2007_2027US_b_18s2il11mcn_10t

Address for correspondence

Rima A Gibbings, rgibbi2@uic.edu, +6308248169.

Can Video Conferencing Facilitate Better Discharge Processes and a Superior Patient Experience?

Jeremy Grummet^a, Nilmini Wickramasinghe^{a,b}, Peter Haddad^{a,b}, Louise O'Connor^a

^aEpworth HealthCare, Melbourne, Victoria, Australia

^bFaculty of Health, Deakin University, Melbourne, Victoria, Australia

Abstract

Video conferencing is emerging as a useful clinical tool. Its application has extended to cover home monitoring, psychotherapy, managing patients with chronic diseases, and other various applications. However, utilizing video conferencing capabilities to facilitate discharge processes has to date been sparsely researched. This study explores the possibilities of video conferencing to facilitate better patient discharge processes at a large Australian not-for-profit tertiary healthcare group.

Keywords:

Hospital Information Systems, Point-of-Care Systems, Telemedicine

Introduction

For patients undergoing surgery in a multi-day admission, a standard care scenario requires that their surgeon will review the patient post-operatively to check on their progress. This is usually done by the specialist attending in person before they may be discharged. However, in the Australian setting, most specialists work at multiple institutions. As a result, reviewing ward rounds, especially of post-operative patients, can be delayed for numerous reasons, which in turn delays management decisions and discharge and may lead to lower patient satisfaction among other problems.

To prevent such delays, an often-used alternative is a "phone round", where the specialist checks their patient by speaking with the patient's nurse by phone. However, a phone round does not allow the specialist to perform the clinically important end-of-the-bed inspection, and does not provide the patient with the reassurance that they have been reviewed by their doctor. Telemedicine, whereby doctors consult via Internet video offers a potential solution. This technology could also be used in the inpatient post-operative setting as an adjunct to the current in-person ward rounds.

This study answers the research question: How can video conferencing facilitate a superior discharge process? The underlying issues to be addressed in this study have arisen from practice and the solution designed focusses on enabling a better patient experience.

Aim

The research aims for this study include:

1. Proof of concept of video conferencing capabilities as an adjunct for post-operative inpatient review.
2. If using video conferencing as an adjunct for post-operative inpatient review will provide improved

patient, staff and surgeon satisfaction, more efficient/effective discharge including 70% of patients discharged by 10 AM, and will provide cost savings.

Methods

A clinical trial will be used to assess the benefits of the developed bespoke telemedicine discharge tool. Specifically, a two arm non-blinded study is designed where 100 patients will experience current standard care for discharge while the 100 patients in the intervention arm will experience standard care and the developed solution. At the completion of the discharge experience all patients will be asked to complete an on-line survey designed and validated to assess patients' level of satisfaction with their discharge experience. Standard statistical techniques will be employed to analyze the survey data.

Doctors (urologists) and nurses (on the chosen wards) make up the clinicians involved in this study. They perform their normal duties regarding discharge in both arms of the study. The only difference being that the intervention arm utilizes discharge with the telemedicine/video conferencing solution. The designed solution will be developed in consultation with a subset of clinicians who will form a focus group, subscribing to a Design Science Research Methodology, which will serve to comment on key design features so that the final solution can then be developed. All clinicians participating in the study will be interviewed to ascertain their level of satisfaction with the developed solution over the standard care approach to discharge. Standard thematic analysis techniques will be employed to analyse the collected interview data.

Conclusion

The proposed use of video conferencing capabilities as an adjunct method for surgeons to conduct inpatient post-operative reviews will lead to a higher patient satisfaction, higher doctor satisfaction, higher ward staff satisfaction, greater efficiency of care delivery, discharge time to 70% by 10 AM, and cost savings.

Address for correspondence

Corresponding author Professor Nilmini Wickramasinghe, Epworth HealthCare & Deakin University.
Email: nilmini.work@gmail.com.

A New Essential Functions Installed DWH in Hospital Information System: Process Mining Techniques and Natural Language Processing

Masayuki Honda, Takehiro Matsumoto

Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences, Japan

Abstract

Several kinds of event log data produced in daily clinical activities have yet to be used for secure and efficient improvement of hospital activities. Data Warehouse systems in Hospital Information Systems used for the analysis of structured data such as disease, lab-tests, and medications, have also shown efficient outcomes. This article is focused on two kinds of essential functions: process mining using log data and non-structured data analysis via Natural Language Processing.

Keywords:

Data Mining, Natural Language Processing

Introduction

Hospital Information Systems (HIS) and Electronic Medical Records (EMR) have been developed for a few decades in Japan. The HIS at Nagasaki University Hospital was updated in January 2015 to include a constant securing of system response. In addition, a secondary usage system of patient information (data warehouse (DWH) system) was built to enable the analysis of structured data including diseases, lab-tests, and medications. [1] Several kinds of event log data produced in daily clinical activities, have not been used for improving hospital activities. Process mining techniques that leverage log data should be introduced into the DWH. One of the problems in the research field of medical informatics is the extraction of knowledge and outcomes from unstructured data. In this article, we focus on the process mining technique "iKnow", which is a Natural Language Processing (NLP) tool developed by InterSystems Corporation.

Methods and Results

Process Mining

By applying process mining techniques to several kinds of event log data derived from HIS we can discover, monitor, and improve clinical processes. Process mining techniques are available in the current HIS [2]. Process mining techniques are necessary in the field of medical information and have been applied to several fields in the HIS [3]. New functions using process mining techniques were constructed and updated in the DWH system in the Nagasaki University Hospital. We used PDCA cycles of the following four steps using the open source framework of "ProM" and the extraction tool "XESame"; (1) Extraction of event log data, (2) Applied process mining techniques, (3) Evaluation of daily clinical processes, (4) Visualization of processes. Several kinds of event log data, such as log-in time and place of system users, pa-

tients moving information, and access point of Wi-Fi data were used.

NLP

Traditional NLP systems are top-down and leverage a thesaurus, ontology, and statistical model. "iKnow", is a bottom-up system that does not require a thesaurus and can automatically index valuable information from the original data independent of the length or semantic complexity of the data. Preliminarily discharge summary data was analyzed; "iKnow" was used to extract data from discharge summaries and the progress notes. Cosine similarity for both entities was calculated. Differences were found in clinical department data as well as in individual patient data.

Conclusions

Process mining techniques can obtain the real-time status of patients. This study aimed to improve activity efficiency, system security, and to provide efficient patient service. The context information provided by "iKnow" can be useful for capturing similar terms or terms with similar meaning. We plan to evaluate the features of "iKnow" and other NLP tools. Lastly, we believe that the quality of medical informatics research will be improved by the findings from this study.

Acknowledgements

Funding and support for this work and other relevant acknowledgements are due to Yasuhiro Kobata and Akio Hashimoto in DataCube Corporation and Benjamin De Boe, Masako Ohira, and Minoru Horita in InterSystems Corporation.

References

- [1] M. Honda, T. Matsumoto, Toward secure Hospital Information System and secondary use system of patient data-based on replacement of HIS in Nagasaki University Hospital *HEC* **123** (2016), 1-9.
- [2] W.M.P. Van der Aalst, Process Mining, Discovery, Conformance and Enhancement of Business Processes, London New York; Springer, 2011.
- [3] A.O. Garcia, Analysis of hospital processes with process mining techniques, *MEDINFO* **2015** (2015), 310-314.

Address for correspondence

Masayuki Honda, Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences
m-honda@nagasaki-u.ac.jp

A Study of Evaluation and Automation Instruction of Management by Law of Community Health Service in China

Hongpu Hu^a, Yue Zhang^b, Yao He^a, Fan Yu^a, Minjiang Guo^a

^a Institute of Medical Information, Chinese Academy of Medical Sciences, Beijing, China

^b School of Management, Beijing University of Chinese Medicine, Beijing, China

Abstract

The goal is to explore a concrete mode for the health administrative department to manage community health service. On the basis of existing health laws and regulations, the study synthesized computer technology, set up the automation system of community health service supervision and management, drew up management norm, management procedure, evaluation system, designed corresponding software, and then completed the spot usage. Four changes of community health service management are realized.

Keywords:

Community Health Services; Automation; Software

Introduction

The Community health service (CHS) is the best health care mode which is selected by developed countries through several decades' explosions [1]. At present, CHS business in China is at the stage of system frame construction after launching and expanding pilots [2]. With no or little uniformity in the laws governing CHS and how CHS is managed, the goal of this study was to explore a concrete mode for the health administrative department to manage CHS.

Methods

First, research the relevant laws, regulations and theories of CHS, analyze the questions, and combine current management situations to build up the norm of supervision and management. Second, conduct a cross-sectional survey of managers of CHS organizations. Then select and build up an expert's database to conduct the expert consultation and select the indexes to evaluate CHS management. Next, the automation software was designed with 4 components: the administration management system, evaluating system, searching system, and analytical system. The last two systems generate forms automatically on the basis of the first system according to the analytical form designed in advance. The field implementation includes preparing experiment, designing investigation questionnaires, and field application. Field application: Two districts in Beijing are chosen as the trial spots and other two districts as the control ones. This research objects mainly involve 4 district health bureaus, 4 CHS centers and 16 CHS stations.

Results (Planned data analysis)

All investigation questionnaires are inputted and checked by the trained graduate students of Capital Medical University.

Analytical contents include the spot investigation information, the expert consulting results, the expert's basic circumstance description of the Delphi method, coefficient calculation of evaluated system and the spot implementation. Collected data will be analyzed by the SPSS12.0 after the logic examining. Data will be analyzed according to describing methods and data needed to be compared will conduct T test and χ^2 test.

Discussion and Conclusions

On the basis of existing health laws and regulations combined with computer technique, the study will set up the automation management system of CHS supervision for the first time in China, and explore one new administrative mode. By implementing functions of automation management, searching, supervision, analysis, evaluation, four breakthroughs can be realized.

Acknowledgements

This study is supported by Fund Projects: National Science Foundation of China (No. 71273280), National Social Science Foundation of China (No.12BGL111).

References

- [1] Yong Bao (2004). Practical Guidebook of Community Health Service. Shanghai: Shanghai Science and Technology Press.
- [2] Wannian Liang, Ya Dong Wang (2005). The Actuality Investigation of Community Health Service in China. Chinese General Practice. 9 (8), 70.

Address for correspondence

Hong Pu Hu hhp4430140@sina.com

Using a Mobile Device Application to Support Emergency Clinicians in Diagnosing Pulmonary Embolism

Chun Yen Huang^{a, b}, Po Liang Cheng^{a, c}, Po Lun Chang^a

^a Institute of Biomedical Informatics, National YangMing University, Taipei City, Taiwan, China

^b Department of Emergency Medicine, Far Eastern Memorial Hospital, New Taipei City, Taiwan, China

^c Department of Emergency Medicine, Buddhist Tzu Chi Medical Foundation, Dalin Tzu Chi Hospital, Chia-Yi County, Taiwan, China

Abstract

Pulmonary embolism is a difficult disease to diagnose in the emergency department. It may be fatal if the diagnosis is missed. Clinical practice guidelines and textbooks publish pretest diagnostic tools and algorithms that facilitate the diagnosis of pulmonary embolism. We developed and administered a questionnaire that determined that such tools can be difficult to remember to use. We also designed an Android application to facilitate diagnosis of pulmonary embolism and used a questionnaire to evaluate the application.

Keywords:

Pulmonary embolism; Decision support systems, clinical; Mobile applications.

Introduction

The symptoms and signs of pulmonary embolism are non-specific, from asymptomatic to shock or death [3]. The mortality rate of pulmonary embolism without proper treatment is up to 25-30% [2]. But if timely anticoagulant treatment is administered, then the mortality rate can be reduced to 2-8%. Therefore, early diagnosis of pulmonary embolism and treatment is very important [1].

Systematic application of existing algorithms or pretest tools published by guidelines or textbooks, such as Wells' score, Geneva score and PERC rule, can support accurate and timely diagnosis [3; 4]. Because assessment can be complex, time-consuming, or inconsistent, routine usage may be challenging.

We designed an Android application that integrates the symptoms, risk factors, algorithm and pretest tools of pulmonary embolism. Then, we developed a questionnaire intended for emergency physicians to identify barriers in diagnosing pulmonary embolism and evaluate the feasibility of the application designed.

Methods

We designed an Android application using MIT App Inventor 2 beta. The context includes the symptoms, risk factors, algorithm, pretest tools derived from the 2014 ESC Guidelines on the Diagnosis and Management of Acute Pulmonary Embolism, as well as from Tintinalli's Emergency Medicine - A Comprehensive Guide, 8th edition. We developed and administered a 10-item questionnaire using Google Forms. The questionnaires were completed by emergency physicians in Far

Eastern Memorial Hospital (FEMH) in Taiwan after using the application for 2 weeks.

Results

We sent out the questionnaire to 38 emergency physicians and 26 physicians filled it (68% response rate). The main reason for non-responders were that they used iOS and not Android. Most physicians thought pulmonary embolism is difficult to diagnose in the emergency department (77%), most physicians agree that pretest tools are helpful in diagnosis (61%), and difficult to remember (84%). After using the application, most physicians felt the application is easy-to-use (77%), facilitates rapid decisions about ordering computed tomography or pulmonary embolism treatment (64% < 6 minutes). Finally, most physicians agree the application is helpful in diagnosing pulmonary embolism (77%). The application will be implemented in a medical center and nurse practitioners will help physicians to use the application for selected patients. Further evaluation of the usability of the application in a real emergency department is in progress.

Conclusion

Timely and accurate diagnosis of pulmonary embolism is clinically significant. The application we designed appears to be user-friendly and acceptable by emergency physicians. It will be piloted in the FEMH emergency department.

References

- [1] S.Z. Goldhaber, L. Visani, and M. De Rosa, Acute pulmonary embolism: clinical outcomes in the ICOPER, *Lancet* **353** (1999), 1386-1389.
- [2] K.T. Horlander, D.M. Mannino, and K.V. Leeper, Pulmonary embolism mortality in the United States, *Arch Intern Med* **163** (2003), 1711-1717.
- [3] S.V. Konstantinides, A. Torbicki, G. Agnelli et al., 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism, *Eur Heart J* **35** (2014), 3033-3069.
- [4] J. Tintinalli, Tintinalli's Emergency Medicine - A Comprehensive Guide, 8th ed., (2015), 388-399.

Address for correspondence:

Chun Yen Huang, e-mail preferred: meureka@gmail.com, +886958035787. 1F., No.56, Aly. 17, Ln. 377, Zhongping Rd., Xinzhuang Dist., New Taipei City 242, Taiwan.

The Impact of PACS on Radiograph Interpretations in an Orthopedic Outpatient Clinic

Petter Hurlen^{a,c}, Arne Borthne^{a,b}, Pål Gulbrandsen^{b,c}

^a Department of Diagnostic Imaging, Akershus University Hospital, Lørenskog, Norway

^b Faculty Division Akershus University Hospital, University of Oslo, Oslo, Norway

^c Helse Sør-Øst Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway

Abstract

We found no significant change in the accuracy of clinicians' reading of skeletal radiographs before and after PACS- introduction. The level of inter-rater agreement between clinicians and radiologists was high in both periods, but they disagreed in significantly more cases after PACS.

Keywords:

Radiology Information Systems, Diagnostic Errors, Program Evaluation

Introduction

It is suggested that diagnostic accuracy may be maintained when images are viewed on monitors rather than on film under optimal conditions [1], [2]. A previous study of chest radiograph interpretation before and after the introduction of a Picture Archiving and Communication system (PACS) indicates that this may apply also in a "real life situation" [3].

This study assesses the impact of PACS on the accuracy of clinicians' radiograph interpretation at an orthopedic emergency outpatient clinic. The study was approved by the Norwegian Social Science Data Service (NSD).

Methods

The basic design of this study was a before-after study using two cross-sectional data collections. Clinicians' and radiologists' assessment of the same images were classified independently as either positive (certain or possible skeletal injury) or negative (all others). Cases where the clinician and radiologist disagreed were independently reviewed by two radiology specialists. Cases with agreement and the results of the review constituted the gold standard.

Differences were analyzed using the independent samples T-test, inter-rater agreement using Cohen's kappa. Significance levels (predetermined at $\alpha < 0.05$) are reported.

Results

Table 1 presents the clinicians' interpretation of the radiographs before and after PACS. There was no significant difference between pre-PACS (0.95) and post-PACS accuracy (0.93, $p=0.36$).

Table 2 presents the classification of the clinicians' interpretation relative to the radiologists'. The level of inter-rater agreement was high in both periods. The Cohen's kappa value was 0.83 before and 0.77 after PACS. However, the clinicians and radiologists disagreed in significantly less cases before than after PACS (7.2% vs 10.2%, $p<0.05$).

Table 1

Pre-PACS		Gold standard	
		positive	negative
Clinician	positive	163 (28,0%)	23 (3,9%)
	negative	8 (1,4%)	389 (66,7%)
Post-PACS		Gold standard	
		positive	negative
Clinician	positive	237 (30,9%)	37 (4,8%)
	negative	13 (1,7%)	481 (62,6%)

Table 2

Pre-PACS		Radiologist	
		positive	negative
Clinician	positive	160 (27,4%)	26 (4,5%)
	negative	16 (2,7%)	381 (65,4%)
Post-PACS		Radiologist	
		positive	negative
Clinician	positive	225 (29,3%)	49 (6,4%)
	negative	30 (3,9%)	464 (60,4%)

Conclusion

Our results indicate that diagnostic accuracy is maintained when film-based radiology routines are replaced by PACS in an emergency orthopedic outpatient clinic.

References

- [1] H. L. Kundel, M. Polansky, M. K. Dalinka, R. H. Choplin, W. B. Gefter, J. B. Kneeland, and W. T. Miller, "Reliability of soft-copy versus hard-copy interpretation of emergency department radiographs: a prototype study," *AJR Am J Roentgenol*, vol. 177, no. 3, pp. 525–528, Sep. 2001.
- [2] G. C. Weatherburn, D. Ridout, N. H. Strickland, P. Robins, C. M. Glas-tonbury, W. Curati, C. Harvey, and C. Shadbolt, "A comparison of conventional film, CR hard copy and PACS soft copy images of the chest: analyses of ROC curves and inter-observer agreement," *Eur J Radiol*, vol. 47, no. 3, pp. 206–214, Sep. 2003.
- [3] P. Hurlen, A. Borthne, F. A. Dahl, T. Ostbye, and P. Gulbrandsen, "Does PACS improve diagnostic accuracy in chest radiograph interpretations in clinical practice?" *Eur J Radiol*, vol. 81, no. 1, pp. 173–177, Jan. 2012.

Address for correspondence

Dr. Petter Hurlen

e-mail: petter@hurlen.no

Comparing Inpatient Falls Guidelines to Develop an ICNP®-Based Nursing Catalogue for ENRs

Miran Jung^a, Insook Cho^b, Jisun Chae^a, Jihye Kim^a, Yeunhee Kim^a

^a Asan Medical Center, Seoul, Republic of Korea

^b Nursing Department, Inha University, Incheon, Republic of Korea

Abstract

This study compared and synthesized clinical guidelines for inpatient falls with the aim of developing an ICNP®-based nursing catalogue for use in electronic nursing records. The identified content will influence nursing services provided to patients and the associated documentation.

Keywords:

Nursing Records; Accidental Falls; Patient Safety

Introduction

The rapid adoption of electronic health records means that electronic nursing records (ENRs) could provide data for reporting nursing quality metrics. Content design issues for ENRs are important for determining how nursing services are represented, recorded, and aggregated. Korea adopted the ICNP® (International Classification for Nursing Practice) early in 2000 as a national standard. Our research team planned to develop an ICNP® catalogue for the patient safety domain, specifically for inpatient falls. The Patient Safety Act established nursing care laws that limits the bedside visits of family members and private caregivers. This situation has led to inpatient falls becoming a high-priority need area. To develop an appropriate catalogue, we reviewed and compared the current evidence in order to identify key nursing components. The presented findings can be used to develop an inpatient fall ICNP® catalogue for use in ENRs.

Methods

Inpatient falls are governed by complex factors, but there is a considerable body of literature on fall prevention and reduction. We assembled the seven resources and toolkits suggested by the Joint Commission with the knowledge gained through research and quality improvement initiatives from six leading organizations. The practice guideline developed by the Korean Hospital Nurses Association and a local guideline of Asan Medical Center were also included. The international classification of patient safety framework [1] was used as a theoretical basis and consensus-based content analysis was employed.

Results

We identified 11 key care components and 98 relevant concepts in 3 categories (Table 1). Several gaps between the two Korean guidelines and other guidelines were identified.

Table 1 - Identified care components and relevant concepts

Category	Key care components	Number of concepts & examples
Patient characteristics	Physical	22 (e.g., age, diagnosis, gait, dizziness, INR)
	Behavioral	8 (e.g., agitation, irritability, delirium, dementia)
	Social	4 (e.g., fall history, drinking, smoking)
Contributing factors/hazards	Environmental risk	9 (e.g., transfer, equipment, lighting, footwear)
	Therapeutic risk	7 (e.g., medications, Foley catheter, operation, IV line)
Actions taken to reduce risk (interventions)	Universal	13 (e.g., toileting, education)
	Impaired mobility	9 (e.g., walking aids, safety aids, helmet, hip protector)
	Behavioral	3 (e.g., triggers of agitation)
	Observation & surveillance	7 (e.g., regular rounding, alarm, pain management)
	Communication	9 (e.g., visual indicator)
	Environmental	7 (e.g., nonslip mat, floor)

Discussion

The available resources and toolkits were found to have a high degree of commonality with each other. However the Korean guidelines had no addresses on communication requirements between clinical disciplines.

Conclusion

The application of both top-down and bottom-up approaches has identified comprehensive evidence-based care components and core concepts for use in nursing statements for ENRs.

Acknowledgements

This study was supported by a grant of the Korea Healthcare Technology R&D project, Ministry for Health, Welfare & Family Affairs, Republic of Korea (No. HI15C1089).

References

- [1] The World Alliance for Patient Safety Group, et al. Towards an International Classification for Patient Safety: the conceptual framework. *Int J Qual Health Care* 21(1), (2009):2-8

Address for correspondence

Insook Cho

email: insook.cho@inha.ac.kr

Willingness to Pay for Elderly Tele-Care System Using Digital Terrestrial Broadcasting

Ogasawara Katsuhiko^{a,b}, Kaga Satoshi^a, Suzuki Teppei^b, Tani Yuji^{b,c},
Higashikawa Fumihiko^d, Enomoto Hisashi^e, Moriyama Hiroyuki^f, Kise Nobuyuki^e

^a Graduate school of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan,

^b Faculty of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan,

^c Asahikawa Medical University Hospital, Asahikawa, Hokkaido, Japan

^d TV Asahi Corporation, Minato, Tokyo, Japan

^e Iwamizawa City, Iwamizawa, Hokkaido, Japan

^f Hamanatsu Information CO., LTD, Iwamizawa, Hokkaido, Japan

Abstract

The aim of this research is to quantify individual willingness-to-pay (WTP) to test effectiveness of Digital terrestrial broadcasting elderly tele-care system. We used contingent valuation method (CVM) to estimate the WTP for this system among 400 citizens living in Japan. The median WTP for this services's monthly fee is estimated to be 431 JPY. The finding suggests that people who cares their health were pay more to use this system.

Keywords:

tele-care system, digital terrestrial broadcasting, WTP
:(willingness to pay)

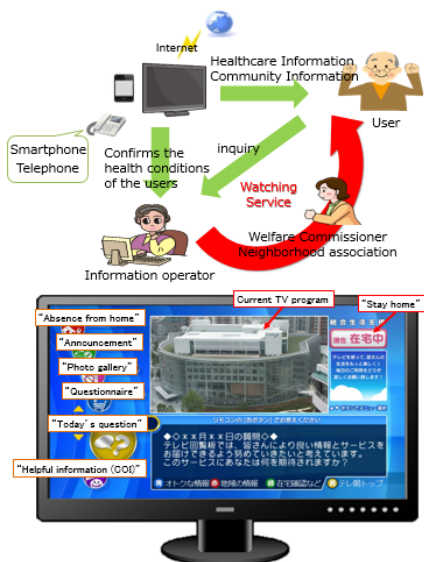


Figure 1 - Elderly tele-care system using digital terrestrial broadcasting

Introduction

In Japan, over the past few years, attention has been increasingly focused on solitary death in relation to the aging society and the

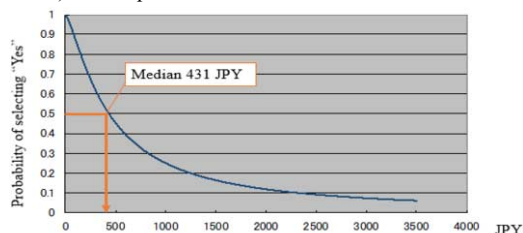
trend towards the nuclear family. Hokkaido University, TV-Asahi, and Iwamizawa-City have implemented a prevention measure with Digital terrestrial broadcasting elderly tele-care system. Although this system is provided for free of charge as a demonstration test, determining the appropriate price is required for Sustainable operation of the service. The aim of this research is to quantify individual willingness-to-pay (WTP) to test effectiveness of Digital terrestrial broadcasting elderly tele-care system.

Methods

We used contingent valuation method (CVM) to estimate the WTP for this system among 400 citizens living in Japan. A questionnaire survey was conducted for people from age 18 to 100 according to Japanese age distribution. To elicit WTP we adopt a "double-bound dichotomous choice method" for asking the respondents whether they are agree or against the price we offered.

Results

The median WTP for this services's monthly fee is estimated to be 431 JPY. The finding suggests that gender (0.66, $p=0.014$), health consciousness (1.08, $p=0.013$), willingness to use (2.38, $p<0.000$), and seeing others less than once a week (1.00, $p=0.057$) made a positive effect on WTP.



$$Pr[Yes] = \frac{1}{1 + \exp\{-5.31 + 0.66sex + 1.08awareness + 2.38willingness + 1.00(only - 1.31ln(hud))\}}$$

Address for correspondence

Katsuhiko Ogasawara

Mail: oga@hs.hokudai.ac.jp Tel: +81-11-706-3409

Address: N12-W5, Kita-ku, Sapporo 060-0812, Japan.

Development of a Quick SOFA-Based Sepsis Clinical Decision Support System in a Tertiary Hospital Emergency Department

Youn-Jung Kim^a, Jae-Ho Lee^{a,b}, Sang-Wook Lee^c, Ming-Young Jung^b, Won Young Kim^a

^a Department of Emergency Medicine, University of Ulsan College of Medicine, Seoul, Republic of Korea,

^b Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic of Korea,

^c Medical Information Office, Asan Medical Center, Seoul, Republic of Korea

Abstract

New definition of sepsis introduced quick sequential organ failure assessment (qSOFA) score. A qSOFA-based sepsis decision support system was designed and developed in an emergency department. The system has functions of automatically retrieving qSOFA score, 3 and 6 hours of sepsis and septic shock treatment bundle, semi-automatic calculation of SOFA score, etc. Early usage and user requests are needed to find aspects of improvement.

Keywords:

Sepsis, Emergency Medicine, Decision Support System

Introduction

The quick sequential organ failure assessment (qSOFA) was introduced as a simple tool to suspect sepsis in emergency departments (ED) by new definition of sepsis and septic shock [1]. The qSOFA is simple and easily retrieved from electronic medical records (EMRs) of ED. To help ED physicians detecting sepsis early and guiding sepsis diagnosis and treatment, a sepsis clinical decision support system (CDSS) was designed and developed in a tertiary hospital ED.

Methods

The research team was composed of ED physicians, a medical informatician, a programmer, and a registered nurse. New sepsis guidelines were reviewed and items which could be included in the sepsis CDSS were selected. Data were retrieved from EMRs and semi-automatic calculation of SOFA scores (different from qSOFA) was performed. This study was performed from March to June, 2016.

Results

A qSOFA-based sepsis template and CDSS were implemented to the EMR of ED Jun 2016. qSOFA score of ≥ 2 was retrieved from nursing information note at ED admission by automatically calculating variables of mental change, heart rate, and systolic blood pressure. Patients with qSOFA score of ≥ 2 score at ED admission were marked as "S" at the end of their names in the list of EMR. Confirming 'suspected infection' process was added to the sepsis CDSS template (Figure 1). The template contained a patient list of qSOFA score of ≥ 2 , initial vital signs, entry field of suspected infection foci, 3 hours and 6 hours of the sepsis and septic shock bundle, semi-

automatic calculation of SOFA score, and dispositions. Data and contents were reviewed and corrected.



Figure 1- A screenshot of the sepsis CDSS template. Translation to English was provided for buttons and text.

Discussion

Active alert system of new suspected sepsis case could be an alternative to adding "S" to end of patient name. ED physicians had to check "Infection or not" because there was no way to identify such information automatically at the time of their visits. SOFA score calculation support was done semi-automatically due to its complexity comparing qSOFA. Initial qSOFA variables were used to calculate qSOFA score because continuous monitoring of vital signs could use large resources of the EMR system.

Conclusion

The qSOFA-based sepsis CDSS was developed to help ED physicians. Researches of early usage and user requests are needed as a next step.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2015R1D1A1A01057091).

References

- [1] Quick sequential organ failure assessment, available at <http://www.qsofa.org/>

Address for correspondence

Jae-Ho Lee, rufiji@gmail.com, +82-2-3010-3350

Development of Safety and Usability Guideline for Hospital Information System

Yura Lee^a, Min-Young Jung^a, Gee Won Shin^b, Sangwoo Bahn^c, Taezoon Park^d, Jae-Ho Lee^{a, e}

^a Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic of Korea

^b Industrial Engineering, Human Interface System Lab, Seoul National University, Seoul, Republic of Korea

^c Industrial and Management Engineering, Myongji University, Seoul, Republic of Korea

^d Department of Industrial & Information Systems Engineering, Soongsil University, Seoul, Republic of Korea

^e Department of Emergency Medicine, University of Ulsan College of Medicine, Seoul, Republic of Korea

Abstract

Hospital Information systems (HIS) provide efficiency in clinical practice, and give users the chances of systemic feedback, recommendations based on evidences. However, HIS without consideration of usability or patient safety could make adverse events. In this study, we suggest safety/usability guideline of HIS, extracted via systemic review of related articles. We categorized the guidelines and recommendations by clinical workflow.

Keywords:

Hospital Information Systems, Patient Safety, Workflow

Introduction

The implementation of hospital information systems (HIS) is known to improve patient safety and quality of care. However, the lack of consideration in ergonomic aspects could result in new problems that could compromise patient safety. The Joint Commission in the United States presented 120 HIT-related errors and their types in the 'Sentinel Event Alert' report in March 2015, titled 'Safe Use of HIT'. In this report, the usability of HIT was pointed out to have correlation with patient safety. Though the HIS in Korea were highly adopted and even exported to other countries, there is no inherent guideline or recommendation for the safety or usability of HIS. Therefore, the guideline for safety and usability is required for the design, maintenance and improvement of HIS.

Methods

The search was conducted based on English literature published since 2000 through the literature search database (PubMed, Cochrane, Embase, Web of Science, CHINAL). The search terms were classified into three categories: Guideline, Healthcare information technology, and Safety / usability. Two researchers independently reviewed the article title and abstract. In case no agreement was reached consensus, the decision was made through consultation with the third researcher or the advisory group.

Results

A total of 7,411 articles were identified through the literature search, and 6,222 documents without duplication were reviewed. 81 documents were selected for full-text review.

Based on the final selection of 14 eligible documents, 402 usability / safety guideline bases were extracted and listed. Extracted guidelines were reclassified according to the workflow of clinical practice. That included the 'Display & Identification' for the user's information perception and selection, the 'Action' of the user's work behavior, the 'Feedback' of the alarms provided by the system after the user's work behavior, 'Communication' corresponding to the exchange of information between the medical personnel or system, and the 'Management' for update of the information required for system operation, user monitoring and maintenance.

Conclusions

By reclassifying the existing guidelines according to the workflow of the user's clinical practice, rather than the existing conceptual and system-based classification, it is expected to be more effectively used to improve the usability and safety of HIS in actual clinical work.

Acknowledgements

This study was supported by a grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea. (HI15C1101).

References

- [1] A complimentary publication of The Joint Commission: Sentinel Event Alert, Issue 54, March 31, 2015
- [2] Chae YM, Going Abroad of Korean Health Information Systems Health Inform Res. 2014 Jul; 20(3): 161–162.
- [3] Chae YM, Yoo KB, Kim ES, Chae H. The adoption of electronic medical records and decision support systems in Korea. Health Inform Res. 2011;17(3):172–177.

Address for correspondence

Corresponding author: Jae-Ho Lee, Department of Emergency Medicine, University of Ulsan College of Medicine, Seoul 138-736, Korea.

Email: jaeholee@amc.seoul.kr

Tel.: +82-2-3010-3350, Fax: +82-2-2045-4126

The Role of Technology in Medication Safety Incidents: Interpretative Analysis of Patient Safety Incidents Data

Valentina Lichtner^a, David Gerrett^b, Ann Slee^c, Noreen Gul^b, Tony Cornford^d

^a Leeds University Business School, University of Leeds, Leeds, UK,

^b NHS Improvement, London, UK

^c NHS England, London, UK

^d Department of Management, London School of Economics, London, UK

Abstract

This is a study of medication safety incidents reported to the NHS in England (UK) associated with the use of digital technology. An interpretative analysis of 888 incidents reports offers insight into uses and features of this technology associated with medication errors and potential patient harm.

Keywords:

Patient safety; Drug therapy, Computer-assisted.

Introduction

Electronic medicines management systems (e.g. Computerised Provider Entry Systems - CPOE) are increasingly used worldwide, giving rise to availability of related patient safety data. A number of studies have provided classifications of the types of errors with this technology and contributing factors. Schiff et al [3], for example, queried incidents reported to the United States Pharmacopeia; more than 63 thousands were classed as related to CPOE. Test case scenarios were generated and used to assess vulnerabilities in current systems. CPOEs were found to lack adequate barriers to protect against wrong orders, or their design made data entry error-prone. In Australia, a study of medication incidents in primary care [1] found IT impacted on patient care, including harm or near misses, disrupted clinical workflow, created inefficiencies and user frustration. Although for some incidents risks had always been present, others were more likely to occur with IT, and some were 'unique to IT' [1]. We undertook a sociotechnical analysis of patient safety incidents reported to the National Reporting and Learning System (NRLS) in England and Wales [2] to better understand the role and impact of digital systems on medication safety in the English National Health Service (NHS). The NRLS contains voluntary anonymised reports from all areas of healthcare.

Methods

The aim was to carry out a qualitative analysis of NRLS data to investigate: the role of the technology in medication incidents; issues with the design or implementation of the technology that are incident-prone; if there are specific types of medications involved. The analysis was informed by a set of principles and theoretical assumptions, as follows:

1. Safety is an emerging property of systems; errors are 'normal' in complex systems.

2. Technology influences the risk of errors occurring. Technology can be made 'safer by design'. Through forcing functions or effective display of information, some human errors can be prevented.
3. Human factors principles and heuristics can aid in the design, evaluation, and configuration of systems.

Data were extracted from the NRLS, to retrieve reports of incidents related to the use of a range of digital systems for the supply and use of medicines over five years. It included all incidents reported in the period from 1/1/2012 to 31/12/2015 (based on date of incident), irrespective of associated degree of harm, incident type and care setting of occurrence. A total of 24,889 records were retrieved. A sample of 888 (approximately 4%) were analysed and coded. Thematic extraction (by VL) was based on initial opportunistic identification and follow up scoping to produce a spectrum of issues until saturation was reached. Content validity was achieved by inspection and ratification by remaining authors. The sample led to identification of 50 themes, and 6 overarching categories: Design, Data, Software, Hardware & Network, Use and Mix-ups. IT related medication safety incidents were across hospital, primary care and community settings, involving a number of IT systems, beyond CPOE. In one case, related to Hardware & Network infrastructure, patient harm was reported as severe,

Acknowledgements

The study is a collaboration with NHS Improvement, funded by University of Leeds EPSRC Impact Acceleration Account.

References

- [1] F. Magrabi, S.T. Liaw, D. Arachi, W. Runciman, E. Coiera, and M.R. Kidd, Identifying patient safety problems associated with information technology in general practice: an analysis of incident reports, *BMJ Qual Saf* (2015).
- [2] NHS England, Patient Safety, 2016. Available from : <http://www.nrls.nhs.uk/>
- [3] G.D. Schiff, M.G. Amato, T. Eguale, J.J. Boehne, A. Wright, R. Koppel, A.H. Rashidee, R.B. Elson, D.L. Whitney, T.-T. Thach, D.W. Bates, and A.C. Seger, Computerised physician order entry-related medication errors: analysis of reported errors and vulnerability testing of current systems, *BMJ Qual Saf* (2015).

Address for correspondence:

Valentina Lichtner, v.lichtner@leeds.ac.uk

Using Modified-ISS Model to Evaluate Medication Administration Safety During Bar Code Medication Administration Implementation in Taiwan Regional Teaching Hospital

Pei-Luen Ma^{ab}, Yan-Wun Jheng^{ac}, Bi-Wei Jheng^d, I-Ching Hou^a

^a School of Nursing, National Yang Ming University, Taipei, Taiwan, China

^b Department of Anesthesia, Sijhih Cathay General Hospital, New Taipei City, Taiwan, China

^c Department of Nursing, Taipei Veterans General Hospital, Taipei, Taiwan, China

^d Department of Nursing, Taoyuan General Hospital, Ministry of Health and Welfare, Taoyuan, Taiwan, China

Abstract

Bar code medication administration (BCMA) could reduce medical errors and promote patient safety. This research uses modified information systems success model (M-ISS model) to evaluate nurses' acceptance to BCMA. The result showed moderate correlation between medication administration safety (MAS) to system quality, information quality, service quality, user satisfaction, and limited satisfaction.

Keywords:

Information systems success model, BCMA, Medication administration safety

Introduction

Medication administration safety (MAS) is the first priority for clinical nurses and one of the BCMA implantation's purposes is to help the clinical nurses reducing the medication error. Therefore, we used the M-ISS model[1] with MAS facet added (e.g. modified-ISS model) to discuss the relationship between MAS to system quality, information quality, and service quality in one Taiwan regional teaching hospital with 862 beds after the BCMA implementation in 2015 (Figure 1).



Figure 1: BCMA interface

Methods

In October, 2016, the questionnaire contained demography, M-ISS model questions with five-point Likert scales, open questions, and one question of the overall satisfied score with 0 to 100 was simple randomly distributed to the nurses who work on the jobs over three months. The CVI was 0.95 and the pilot testing's Cronbach's α , with 30 samples, was 0.932. The statistical methods included descriptive and inferential statistics for quantitative data and content analysis for open questions.

Results

We received 350 questionnaires with 96.6% effective response rate. The descriptive statistic result showed 96.8% participants were females, the mean age was 31.7 with a bachelor degree, and the consciousness of computer ability was between good and average. The MAS was moderate correlation to system quality, information quality service quality, and user satisfaction (Figure 2). After analyzing the open questions, the advantages of BCMA include reducing the medical administration time, quick inquire the medication information and identify easily. The disadvantages include slow connection, poor sensitivity of the bar code sensor, and system crashes.

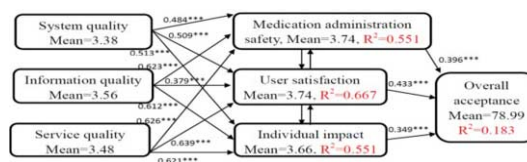


Figure 2: The correlation of each M-ISS model facet

Conclusions

This study shows MAS was affected by system quality, information quality, service quality, and user satisfaction, which has good explanatory power. According to the open questions, MAS will be promoted by renewing the medication data in real time, and elevating the system's efficacy

Acknowledgements

Funding sources for the work was supported by the Taoyuan General Hospital, Ministry of Health and Welfare.

References

- [1] W. H. DeLone and E. R. McLean, The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, *Journal of Management Information System* **19** (2003), 9-30.

Address for correspondence

Corresponding Author: I-Ching Hou
Address: Room 407, School of Nursing, No.155, Sec. 2, Linong St., Beitou Dist., Taipei City 112, Taiwan; E-mail
Address: evita@ym.edu.tw; Phone: 866-2-28267315

Decision Support System for Medical Care Quality Assessment Based on Health Records Analysis in Russia

Taranik Maksim^a, Georgy Kopanitsa^{a,b}

^a Tomsk Polytechnic University, Tomsk, Russia,

^b Tomsk State University for Architecture and Building, Tomsk, Russia

Abstract

The paper presents developed decision system, oriented for healthcare providers. The system allows healthcare providers to detect and decrease nonconformities in health records and forecast the sum of insurance payments taking into account nonconformities. The components are ISO13606, fuzzy logic and case-based reasoning concept. The result of system implementation allowed to 10% increase insurance payments for healthcare provider.

Keywords:

Standards, Case Based Reasoning, Quality Assurance

Introduction

Medical care process in Russia was unified [1] and medical standards were implemented. Standard describes the necessary medical services and drug prescriptions for groups of diagnoses. Medical services contain coefficients, which determine the probability of application. This fact identifies medical standard as reference document with uncertainty and makes fuzzy logic [2] as a relevant tool to work with them. Medical organizations (MO) have to provide their activity with standards. Insurance companies (IC) makes evaluation for MO's medical care by inspecting health records. On the result of evaluation, IC makes an invoice document that contains data of insurance payment for a medical service. However, this evaluation can be based not only on standards, but also on medical expert's personal experience. Expert produces a document, which presents all nonconformities. This document evaluates the quality of medical care specifies the sum of insurance payments for the analyzed treatment case. Medical care quality assessment based on health records evaluation by expert. Expert can evaluate using medical standards (explicit knowledge) and personal experience (implicit knowledge). In the Russian practice, the use of implicit knowledge is more common. We propose a concept of a decision support system for medical aid quality management, which allows healthcare providers to detect and reduce medical nonconformities in health records taking into account uncertainty and forecast the sum of insurance payments.

Methods

The object of the research is planned surgical treatment in the Institute of Microsurgery (Tomsk, Russia). We acquired a set of health records of 2015 inspected by medical experts and the results and conclusions of their evaluation. Combinations of treatment cases of 2015 and results of expert evaluation determined a training set. Linking the medical conclusions to the results of the expert evaluation allowed us to formulate the logical inference rules. We used the Euclidean metric to limit the set of cases and Mamdani algorithm to determine the probability coefficients. Using developed models, we determined how many

medical documents are required at each step of the medical care process. Mamdani algorithm was used to evaluate the quality of treatment case and to forecast insurance payments based on the detected nonconformities.

Results

The system implementation in Institute of Microsurgery's treatment process in the beginning of 2016 would allow to significant decrease the total sum of deduction from insurance payments (table 1) by detecting nonconformities in health records and insurance payments forecasting.

Table 1 –Results before and after system implementation

Month	Fine sum before implementation	Fine sum after implementation
January	26359.853	10016.74
February	10772.973	7002.43
March	15904.979	8111.54
April	24854.181	10190.21
Mai	8906.788	1068.81
June	18725.46	0.00
July	2990.136	508.32
August	7305.686	3068.39
September	14187.242	7093.62
October	9818.673	0.00
November	13678.28	6018.44
December	23486.35	6106.45

The key results of intellectual system implementation:

- The possibility of the insurance payments forecasting;
- The insurance payments for medical care were increased by correction of health records;

Conclusion

The methods presented this research has demonstrated an effectiveness on insurance payments forecasting. Presented concept can be also applied to others solution in public health, hybrid inference method based on fuzzy logic and CBR can be used for different relevant problems, which require decision-making process.

References

- [1] M.A.Taranik, G.D. Kopanitsa. Analysis of the process of control quality of medical service in the scope of compulsory health insurance program, Journal of control and computer science, Tomsk state university, 3 (32) (2015), 75–85.
- [2] L.A. Zadeh. Fuzzy Sets, Information and Control (1965), 8, 338–353.

Address for correspondence

Georgy.kopanitsa@gmail.com (Georgy Kopanitsa)

Method to Identify Diagnostic Rules for Pancreatic Cancer Using Laboratory Data Based on Bayesian Estimation

Shirou Manabe^a, Yoshie Shimai^a, Taizo Murata^b, Ayumi Fujii^b, Kanayo Ueda^b, Akito Nakagawa^a, Toshihiro Takeda^a, Naoki Mihara^a, Yasushi Matsamura^a

^a Department of Medical Informatics, Osaka University Graduate School of Medicine, Osaka, Japan

^b Department of Medical Informatics, Osaka University Hospital, Osaka, Japan

Abstract

Early diagnosis and treatment of pancreatic cancer is challenging. We attempted to find diagnostic rules for pancreatic cancer from laboratory data in the Osaka University Hospital's data warehouse using Bayesian estimation. We calculated the pretest odds based on the number of laboratory tests and the cutoff value at which the diagnostic accuracy is over 20%. By this method, we identified diagnostic rules of 6 types for one item and 79 types for 2 items. Pancreatic cancer is difficult to detect from only general laboratory tests. However, this method may be promising in early diagnosis.

Keywords:

Bayes Theorem; Pancreatic Neoplasms; Chemistry, Clinical

Introduction

To treat pancreatic cancer, it is most effective to find it early and start treatment. However early detection of pancreatic cancer is difficult. Earlier diagnosis may be achieved if guidelines will be developed from accumulated test data and electronic health record systems to alert doctors when cancer is suspected. Research for pancreatic cancer using pathological data has been reported. We tried to identify such diagnostic guidelines by Bayesian estimation using both laboratory test and cancer registration data.

Methods

During a set time period, the pretest odds for each cancer laboratory test can be calculated from the numbers of the tests of cancer and non-cancer patients. When two kinds of tests are measured at the same time, the pretest odds can be similarly calculated from the number of the test where these two items are measured at the same time.

Bayes' theorem is expressed as follows. (LR^+ ; a likelihood ratio for a positive test result)

$$\text{Posttest odds} = \text{pretest odds} \times LR^+$$

By dividing posttest odds by pretest odds, the target positive likelihood ratio for obtaining posttest odds is obtained. When the sensitivity is 0.1 or higher, the positive likelihood ratio at which the posttest odds are 0.25 or higher was adopted as the target positive likelihood ratio. The cut-off value that exceeds the target positive likelihood ratio was defined as a diagnostic rule of pancreatic cancer in one laboratory test item.

when two independent laboratory test items are performed consecutively, the posttest odds can be obtained by the following formula using the Naive Bayes model. ($LR1^+$, $LR2^+$ are positive likelihood ratios of test 1 and test 2)

$$\text{Posttest odds} = \text{pretest odds} \times LR1^+ \times LR2^+$$

However, in most cases, the two laboratory test items were not independent. Therefore, the posttest odds in such case were corrected using the correlation coefficient r .

$$\text{Posttest odds} = \text{pretest odds} \times LR1^+ \times LR2^+ \times (1 - (1 - 2 / (LR1^+ + LR2^+)) \times r)$$

When examining all the combinations between two items, the amount of calculation becomes enormous. Therefore, it was decided to make a combination using five kinds of values, which are the minimum value, the 25% value, the median value, the 75% value, and the cutoff value having the sensitivity of 0.1 for each test item. Of the combinations exceeding the target positive likelihood ratio, the combination with the lowest value was taken as the diagnostic rule for that pair of laboratory test items.

Results of 38 laboratory tests such as biochemistry, blood count, and tumor marker were obtained from 17,202 patients (include 288 pancreatic cancer patients) in January 2013. Their data was used to obtain one or two items' pretest odds and correlation coefficient for each test for pancreatic cancer.

176 pancreatic cancer patients registered from January 2007 through December 2011 were selected as the patient group. And 21,106 patients registrations from the same period for other diseases were defined as a control group. For the 38 laboratory test items, a cut-off value, at which posttest odds is 0.25 or greater and sensitivity is 0.1 or higher for each test, was calculated from the maximum value within a month prior to diagnosis.

This study was approved by the Institutional Review Board at Osaka University Hospital.

Results

Six types of diagnostic rules (e.g. ALP, r-GTP) were identified for one laboratory test item alone, and 79 types of diagnostic rules (e.g. CA19-9 x CEA, ALP x Neu) were identified for two laboratory test items.

Conclusions

In clinical decision support systems, timeliness of diagnosis and diagnostic accuracy are important. Using Bayesian estimation appears to be a useful method in facilitating pancreatic cancer diagnosis.

Address for correspondence

Shirou MANABE, Osaka University Graduate School of Medicine, Department of Integrated Medicine, Medical Informatics, 2-15, Yamada-oka, Suita, Osaka 565-0871, JAPAN
manabe@hp-info.med.osaka-u.ac.jp

The Evaluation of the Health Information Exchange with the Number of Usage and the Introduction of Outpatient to the Hospitals at Nagasaki Japan

Takehiro Matsumoto, Masayuki Honda

Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences, Japan

Abstract

Health Information Exchange (HIE) has been gradually spreading in Japan. But the effect and evaluation of these systems have not been proven. The aim of this study is to examine the effect of the Japanese HIE through the evaluation of the clinical activities in Nagasaki University Hospital. The electric medical records of 5335 patients were shared in the Ajisai-net, which was Japanese HIE in Nagasaki, from 2009 to 2015. And the number of shared patients increased every year, and the annual average number was 762.1. On the other hand, the number of the patients who was introduced from clinics to our hospital and whose medical records were shared in the Ajisai-net increased every year. In conclusion, the usage of Japanese HIE is useful for the clinical consultation and medical care, and these systems have the good effect on the hospital management in Japan.

Keywords:

Health Information Exchange; Electronic Health Records

Introduction

Health Information Exchange (HIE) has been gradually spreading in Japan. The Japanese Medical Association reported the number of these networks were at least 245 all over Japan, and were increasing in 2015. Though most network had started recently, the effect or evaluation of these networks were not proved yet. Ajisai-net, one of the most popular HIE, has been used since 2004. The electric medical records of 31 big hospitals of Nagasaki Prefecture are shared among hospitals, clinics, pharmacies and other medical institutions, the number of the medical institutions were over 300. The aim of this study is to examine the effect of the HIE of Japan through the evaluation of the clinical activities in Nagasaki University Hospital which joined Ajisai-net in 2009.

Methods

The electric medical records of 5335 patients were shared with their consent in this Ajisai-net from 2009 to 2015. We evaluated the number of electric medical records of Nagasaki University Hospital in Nagasaki city which were shared in this Ajisai-net from 2009 to 2015. And we compared the rate of introduction to our outpatient clinic after usage of Ajisai-net each year. And we also evaluated the rate of the usage of this system from inside and outside of Nagasaki city in Nagasaki Prefecture.

Results

The number of shared patients increased every year, and the annual average number was 762.1. And the number of last

year, 2015, was 1,018. The annual average number of the usage of this system was 10.7. The average rate of the patients who was introduced from clinics to our hospital was 42.2%, and this rate also increased every year (Table 1). The rate of 2015 was 53.2% which was the largest among study year. On the other hand, the average rate of the usage from outside of Nagasaki city was 11.7%, and this rate had been gradually increasing tendency every year. Of the participating clinics which is inside of Nagasaki city, the average rate of facilities actually used was 72.3% a year.

Table 1 The number and rate of the usage from 2009 to 2015

Year	# M	# F	All	Mean Age	Introduction Rate	Actually Used Facilities (B)	Average (A/B)	
2009	173	187	360	56.7	84	23.3%	24	15.0
2010	312	412	724	58.5	231	31.9%	53	13.7
2011	302	379	681	56.9	252	37.0%	60	11.4
2012	381	375	756	59.1	298	39.4%	79	9.6
2013	398	471	869	59.7	395	45.5%	84	10.3
2014	420	507	927	58.8	447	48.2%	96	9.7
2015	501	517	1,018	58.3	542	53.2%	102	10.0
Sum	2487	2848	5335	58.5	2249	42.2%	498	10.7

Discussion

The number of usage had been increasing every year. And the rate of introduction of the same patient to our hospital had been increasing every year too. And the usage of the clinics outside of Nagasaki city had been increasing. These results show that Ajisai-net is useful for most users, and has incentive effect on the patient introduction. We estimate that massive and high quality medical information and data from large hospitals will be useful for clinical consultation and medical care every day. Furthermore the rate of usage from outside of Nagasaki city had been increasing. This result and the increase of introduction patients show Japanese HIE has the effect on the stabilization of hospital managements.

Conclusion

These results show the usage of Japanese HIE like Ajisai-net in Nagasaki is useful for the clinical consultation and medical care, and also show these systems has the good effect on the hospital management in Japan.

Address for correspondence

Takehiro Matsumoto, Department of Medical Informatics, Nagasaki University Graduate School of Biomedical Sciences
tmatsumo@nagasaki-u.ac.jp

Computerized Cognitive Rehabilitation: Comparing Different Human-Computer Interactions

Silvana Quaglini^a, Anna Alloni^a, Barbara Cattani^b, Silvia Panzarasa^a, Caterina Pistarini^b

^a Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy

^b IRCCS Foundation Hospital Salvatore Maugeri, Pavia, Italy

Abstract

In this work we describe an experiment involving aphasic patients, where the same speech rehabilitation exercise was administered in three different modalities, two of which are computer-based. In particular, one modality exploits the “Makey Makey”, an electronic board which allows interacting with the computer using physical objects.

Keywords:

Computer-Assisted; Aphasia; Rehabilitation.

Introduction

Linguistic communication deficits resulting from brain damage have a strong impact on patients' daily life. The purpose of cognitive rehabilitation (CR) is the return to an acceptable quality of life, which can be obtained by stimulating the language production and comprehension, while improving other cognitive functions. CR, which has been traditionally paper-based, can now be supported by computers. This allows generating exercise sessions characterized by high variety, minimizing the risk of boring patients, and increasing their compliance. It also allows automatic detection of patients' performance, as well as data storage for further analysis. Finally, patients, once trained at the hospital, could continue their rehabilitation plan at home with a telemedicine system. These properties, which in principle could benefit the entire CR process, unfortunately collide with some practical problems, such as the lack of familiarity of many patients with computers. To mitigate this problem, we have experimented a new interface, where the patient interacts with the computer through physical objects. The new modality has been integrated into CoRe, a system for computerized rehabilitation we developed during last years [1].

Methods

We exploited the Makey Makey (MM), an USB electronic board running on top of the Arduino system. It represents an alternative to keyboard and touch-screen, since it allows to “map” any physical conductive object to elements of the keyboard. As a first test, we implemented an exercise for aphasic patients, namely a “complete the word” task. The patient must choose among 6 options (3 consonants and 3 vowels), to complete a word (e.g., APPLE), given its picture and initials (e.g., “AP”). In our trial, three patients with Broca's aphasia, aged 38, 46 and 69 years old respectively, performed several sessions of the exercise in three modes. First, using paper; second, using the touch screen; third, using MM, with the picture and initials shown on the screen while the 6 option letters, large and chromed, were fixed on a mobile lectern (Figure 1). The order of the modes was changed from session to session to avoid bias in subsequent statistical analysis (the patient may decrease its performance in the last minute of the

session because of tiredness). Each session consisted of a set of 20 stimuli, always different, consisting of two-syllable words with their images, to be completed in a 7-minute time limit. Thus, 420 stimuli (7 set x 20 stimuli x 3 modes) were selected avoiding verbs and digraphs and employing high-definition images with neutral background. For each session we measured total running time, accuracy (number of words completed without errors), any aid provided by the therapist, and mistakes. At the end of each session we asked patients to rate the 3 modalities in order of preference.



Figure 1- Paper-based, touchscreen, and makey-makey interaction (wolf=lupo, wave=onda in Italian)

Results

The rehabilitation sessions went smoothly and did not require additional time by the therapist for the preparation of the rehabilitation setting. Performance was quite similar among the 3 modalities, confirming that a different interaction does not affect the result. All patients, in time, have improved execution times, particularly with the PC-based modes. Patients were also positively affected by the unusual unfolding of sessions. All patients preferred the computerized mode to the paper: P1 was indifferent between touch-screen and MM, while P2 and P3 slightly preferred touch-screen.

Discussion and Conclusion

This work showed feasibility and ease of use of PC and MM applications in aphasic patients. However, our three patients, before the injury, had a good familiarity with the PC, and this probably did not allow to fully assess the MM possible benefits. The trial will thus continue with patients with older age and lower computer literacy as well as exercises that involve the use of conductive objects other than chrome letters, such as fruits, vegetables or even graphite drawings.

References

- [1] Alloni A, Sinforiani E, Zucchella C, Sandrini G, Bernini S, Cattani B, Pardell DT, Quaglini S, Pistarini C. Computer-based cognitive rehabilitation: the CoRe system. *Disabil Rehabil.* (2015) Oct 27:1-11

Address for correspondence

Silvana Quaglini. E-mail silvana-quaglini@unipv.it

In Vitro Comparative Study Between Conventional and Computer-Assisted Surgery Methods for Planning and Resection of Bone Sarcomas

L.E. Ritacco^a, L. A. Aponte-Tinao^b, D. L. Muscolo^b, M.A. Ayerza^b, I. Albergo^b,
G.L. Farfalli^b, Risk M^{cd}, F. Gonzalez Bernaldo de Quiros^a, F.E. Milano^{ac}

^a Virtual Planning and Navigation Unit, Department of Health Informatics,
Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

^b Institute of Orthopedics "Carlos E. Ottolenghi", Hospital Italiano de Buenos Aires, Buenos Aires, Argentina.

^c National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina

^d Instituto Tecnológico de Buenos Aires, Buenos Aires, Argentina

Abstract

This poster aims to achieve an "in vitro" comparative study between three methods: 2D digital images planning and execution without navigation (freehand with ruler and caliper), 3D planning and execution without navigation (freehand with ruler and caliper) and 3D planning and execution guided with navigation. 3D planning and navigated procedures potentially improve sarcoma resection.

Keywords:

Informatics, Virtual Navigation, Sarcoma

Introduction

Surgeons plan surgeries using bi-dimensional images from magnetic resonance and tomography to define the tumor extension and then handle standard tools intraoperatively (caliper and ruler) to resect bone sarcoma. We wonder if surgeons have more information about the spatial tumor situation before and during the surgical procedure would it reduce the risk of potential sarcoma recurrence? Thus, the aim of this comparative study is to reach ideal "in vitro" conditions with an experimental design to answer these questions: Which type of planning method is safest for planning an oncologic margin? Which type of executing method is safest for sarcoma resection?

Methods

Two surgeons specialized in bone tumors unfamiliar with computer-assisted techniques and a 2nd year resident were evaluated to measure the accuracy and the impact in bone sarcoma resection according to the tools used. Plastic bones: a proximal femur, a distal femur, a humerus, a pelvis considering iliac wing and acetabulum. The methods evaluated were 2D digital images planning and execution without navigation (freehand with ruler and caliper), 3D simulation scenario planning and execution without navigation (freehand with ruler and caliper) and 3D simulation scenario planning and execution guided with navigation.

Results

Surgeon A, B and C did not remove the tumor when planning in 2D and executing with the freehand method based on those plans. The 3D planned and freehand guided resections exceeded the 3mm threshold a 66% of the times, while the navigated assisted resections a 20% of the times. For the 5mm threshold, the 3D planned and freehand guided resections exceeded it a 41% of the times while the 3D planned and navigated assisted resections a 3% of the time. There is no apparent difference between 3D planned non-assisted

resections and 3D planned assisted resections when evaluating safe margin violations. The navigation-assisted resections are closer to the target resection.



Figure 1 – A Plastic Pelvic bone showing color lines that depict each planning and execution method. Black line: 2D planning and freehand. Red line: 3D planning and freehand. Green line: 3D planning and navigation guidance. The yellow circle shows the estimated location of the tumor.

Conclusions

The 2D planning derived in a wrong resection, leaving tumoral tissue inside the patient. The 3D planning method potentially improved the results. There is no significant difference between 3D planned non-assisted resections and 3D planned navigation-assisted resections. The proposed model is on its experimental stage. The model allows physicians to compare advantages and disadvantages of tools and methods used in oncologic surgeries. 3D planning and navigation are potential assets in order to acquire accuracy and to reach an optimum margin in tumor resections.

References

- [1] Khan FA, Lipman JD, Pearle AD, Boland PJ, Healey JH. Surgical technique: Computer-generated custom jigs improve accuracy of wide resection of bone tumors. Clin Orthop Relat Res. 2013;471(6):2007-2016.

Address for correspondence

Lucas Eduardo Ritacco, MD, PhD Institute of Orthopedics "Carlos E. Ottolenghi" Italian Hospital of Buenos Aires, Potosí 4247 (C1199ACK), Buenos Aires, Argentina
(lucas.ritacco@hospitalitaliano.org.ar). Contact phone: +54 9 11 59766685

Effects of Computerized Guideline-Oriented Clinical Decision Support System on Glycemic Control in Diabetic Patients: A Systematic Review and Meta-Analysis

Ryota Sakurai^a, Kazuhiko Ohe^b

^a Department of Healthcare Information Management, The University of Tokyo Hospital, Tokyo, Japan

^b Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Abstract

In a meta-analysis of 4 trials, computerized guideline-oriented clinical decision support system (CDSS) for healthcare providers showed a similar change in glycated hemoglobin (HbA_{1c}) from baseline to follow up (weighted mean difference (95% confidence interval (CI)): -0.29 (-0.74, 0.16), $p=0.212$), but a reduced proportion of patients with HbA_{1c} $\geq 7.0\%$ at follow up (odds ratio (95%CI): 0.85 (0.74, 0.97), $p=0.014$) compared with the control group, suggesting the need for the development of multifaceted computerized CDSS.

Keywords:

Computers, Guideline, Meta-Analysis

Introduction

A computerized guideline-oriented clinical decision support system (CDSS) for healthcare providers is expected to improve clinical outcomes in patients with chronic diseases, such as diabetes. However, it has not been well characterized whether this type of CDSS is effective for diabetic care. Therefore, we conducted a systematic review and meta-analysis of clinical trials comparing computerized guideline-oriented CDSS with control on glycemic control in diabetes.

Methods

The search terms included “guidelines,” “clinical,” “decision,” “support,” “system,” and “diabetes” in PubMed, the Cochrane Library, and Web of Science. The same terms or relevant studies were also queried on the website of the U.S. National Institute of Health and relevant reviews. The endpoints were a change in glycated hemoglobin (HbA_{1c}) from baseline to follow up and a proportion of patients with HbA_{1c} $\geq 7.0\%$ at follow up. The computerized guideline-oriented CDSS for healthcare providers was defined as that which provides any clinical advice or recommendations automatically generated by computer, according to clinical guidelines.

Results

A total of 4,008 patients (2,037 patients allotted to the intervention group and 1,971 to the control group) in 4 studies (2 randomized controlled trials (RCTs) and 2 cluster RCTs) were included in this study. While the change in HbA_{1c} in the intervention group was not statistically significant compared with the control group (weighted mean difference (95%

confidence interval (CI)): -0.29 (-0.74, 0.16), $p=0.212$), the proportion of patients with HbA_{1c} $\geq 7.0\%$ was significantly lower in the intervention group than in the control group (odds ratio (OR) (95%CI): 0.85 (0.74, 0.97), $p=0.014$).

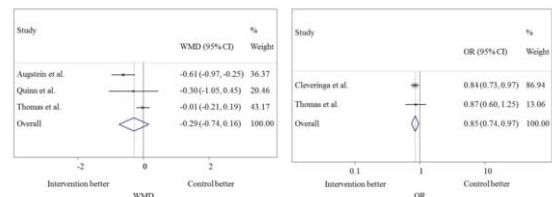


Figure – Forest Plot for the Change in HbA_{1c} (Left) and the Proportion of Patients with HbA_{1c} $\geq 7.0\%$ (Right)

Conclusions

In this meta-analysis, computerized guideline-oriented CDSS for healthcare providers was partially effective on glycemic control in diabetic patients as compared to control. To achieve more improvement of clinical outcomes in diabetic patients, it seems that the development of a multifaceted computerized CDSS, in which not only healthcare providers but also patients could participate and cooperate, would be more effective.

In addition, it might be useful to perform a sensitivity analysis, such as a trial sequential analysis, to adjust for random error risk in a meta-analysis including a small number of trials [1].

Acknowledgments

The authors thank Heidi N. Bonneau for her editorial review. The authors have nothing to disclose regarding this study.

References

- [1] J. Brok, K. Thorlund, C. Gluud, and J. Wetterslev, Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses, *J Clin Epidemiol* **61** (2008), 763-769.

Address for correspondence

Ryota Sakurai, MD, PhD, MPH. Department of Healthcare Information Management, The University of Tokyo Hospital
7-3-1 Hongo, Bunkyo-ku, Tokyo. yoshimune-tyk@umin.ac.jp

Use of an Off-the-Shelf Corporate Information Tool to Track a High-Level-Disinfection Process

^aCarolina Salazar, ^aJaime de los Hoyos, ^aFrancisco Izurieta, ^aDaniel Jara, ^{b,c}Daniel Capurro

^aSalud UC CHRISTUS Healthcare Network, Pontificia Universidad Católica de Chile, Santiago, Chile

^bDepartment of Internal Medicine, School of Medicine, Pontificia Universidad Católica de Chile

^cDepartment of Biomedical and Health Informatics, University of Washington, Seattle, USA

Abstract

High-Level Disinfection (HLD) of endoscopic equipment is a critical process to ensure patient safety. Although commercial tracking software is available, here we demonstrate the successful design and implementation of a HLD tracking system using pre-existing corporate information tools. After a brief development and implementation stage, the HLD unit is successfully tracking the process and are able to create simple visualizations to monitor and improve the existing procedures.

Keywords:

Infection Control, Health Information Systems

Introduction

Endoscopic procedures are first-line diagnostic and therapeutic tools. Given the structure and thermolabile nature of their components, endoscopic equipment cannot go through common sterilization processes; they are considered semi-critical equipment, they must go through a process called High-Level Disinfection (HLD)[1].

HLD is a complex, unidirectional process that requires strict supervision of every step. The main steps in HLD include a) washing and decontamination b) disinfection, and c) packaging and delivery. Complete traceability of this process is paramount.

Our local HLD unit used a paper-based traceability system, which hindered the re-use of the captured information to monitor the process, detect failures and implement quality improvement actions. Although equipment-tracking softwares are available for purchase, we sought to implement one using our already available corporate information tools.

Methods

Our unit documented the current, paper-based HLD process with key users. We modelled the corresponding business facts and structured data needed to support said process Microsoft InfoPath. Once validated, the form was loaded and made available as a Microsoft SharePoint site, and specific user accounts were created. We conducted a week-long testing period, in which users simultaneously documented HLD equipment check-in and check-out in both the paper-based method, and the new SharePoint site. Finally, the HLD process began using the SharePoint site exclusively to support and document their workflow. Designated users can readily extract all records into a Microsoft Excel spreadsheet for review and analysis.

Results

As of December 2016, the HLD SharePoint site has over 2,800 records, and continues to be successfully used to track the HLD process, and our institution was to extract and analyze related information. This enables the creation of quick analyses to make informed decisions based on error rates, turn-around-times, etc. Figure 1 offers a quick visualization of the distribution of equipment admission times to the HLD unit which could inform future decisions on staffing.

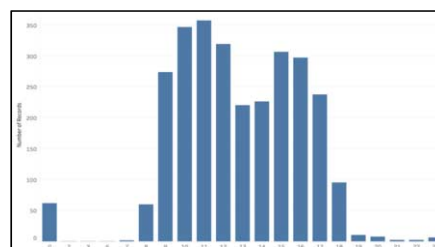


Figure 1 - Distribution of equipment admission times, by hour of the day, during the operation of the new tracking system

Conclusion

This experience demonstrates the practical possibility to use already available corporate information tools to solve practical healthcare information-related problems.

Acknowledgements

We would like to thank the HLD personnel for their collaboration on the HLD process modelling and validation

References

- [1] Guideline for Disinfection and Sterilization in Healthcare Facilities, 2008 [Internet]. Infection Control Practices Advisory Committee (HICPAC), Center for Disease Control and Prevention. [cited December 21, 2016] Available from https://www.cdc.gov/hicpac/pdf/guidelines/Disinfection_Nov_2008.pdf

Address for correspondence

Daniel Capurro, MD, PhD
Diagonal Paraguay 362, 5th Floor
Santiago, Chile. 8330074.

Using Fast Healthcare Interoperability Resources (FHIR) for the Integration of Risk Minimization Systems in Hospitals

Kutaiba Saleh^a, Stephan Stucke^b, Alexandr Uciteli^c, Sebastian Faulbrück-Röhr^b, Juliane Neumann^d, Kais Tahar^c, Danny Ammon^a, Tobias Schmidt^a, Thomas Neumuth^d, Andreas Besting^e, Frank Porthoine^e, Heinrich Herre^c, André Kaeding^b and Martin Specht^a

^aJena University Hospital, Jena, Germany,

^bGMC Systems mbH, Ilmenau, Germany

^cInstitute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig, Germany

^dInnovation Center Computer Assisted Surgery (ICCAS), University of Leipzig, Leipzig, Germany

^eSurgiTAIX AG, Herzogenrath, Germany

Abstract

With the growing strain of medical staff and complexity of patient care, the risk of medical errors increases. In this work we present the use of Fast Healthcare Interoperability Resources (FHIR) as communication standard for the integration of an ontology- and agent-based system to identify risks across medical processes in a clinical environment.

Keywords:

Standards; Decision Support Systems, Clinical; Harm Reduction

Introduction

Current studies provide an estimation of more than 400,000 lethal harms due to preventable medical errors during hospital treatment per year, with 10- to 20-fold serious harms occurring [1]. Clinical decision support systems (CDSS) can assist in preventing such errors, but have not yet integrated the new messaging standard Fast Healthcare Interoperability Resources (HL7 FHIR [2]). In the research project “OntoMedRisk”, we developed an agent-based system to identify risks based on a special ontology [3]. This system aims at supporting medical staff with risk detection and error prevention. A requirement for practical demonstration is the integration of this system into an existing hospital IT infrastructure, which was carried out at Jena University Hospital (JUH).

Methods

Cochlear Implantation (CI) has been chosen as a first exemplary clinical treatment process. To detect risks during CI treatment, relevant medical information from clinical subsystems must be extracted and processed by the system. In our concept, we combined the data retrieval agent of the DSS with a FHIR server to extract data either directly from the clinical subsystems or via communication server to systems which do not yet support FHIR.

Results

We implemented a FHIR server at JUH for medical data collection for risk detection in the CI treatment process, where singular data is extracted and processed in a timely manner.

For communication and storage, certain clinical data key performance indicators relevant for risk detection like age of the patient, cranial thickness, or duration of deafness, were modeled using FHIR resources like “patient”, “observation”, “encounter” and processed by the agent system. Through the use of an ontology, alerts are generated during treatment as a support to the clinical staff. Since FHIR is a new standard, many relevant clinical systems might not be providing a correspondent interface. In these cases, the use of a communication server is a transitional solution.

Conclusions

We obtained as findings that CDSS can collect medical data from existing clinical information systems through the use of modern communication standards like FHIR. The “OntoMedRisk” system relies on real-time access to discrete data generated in different subsystems, and we could provide these as FHIR resources for risk detection and presentation of the results to clinical staff. Further investigation on the usability of this software at JUH is planned. In the future, we plan to extend system functions to other treatment processes.

Acknowledgements

This work has been funded by the BMBF under the project “OntoMedRisk” (grant no. 01IS14022).

References

- [1] J.T. James, A new, evidence-based estimate of patient harms associated with hospital care, *J Patient Saf* 9 (2013), 122-128.
- [2] HL7. FHIR [Internet]. 2016. Available from: <https://www.hl7.org/fhir/>
- [3] A. Uciteli, J. Neumann, K. Tahar, K. Saleh, S. Stucke, S. Faulbrück-Röhr, et al., Risk Identification Ontology (RIO): An ontology for specification and identification of perioperative risks, in: *7th Workshop on Ontologies and Data in Life Sciences*, Halle (Saale), Germany, 2016.

Address for correspondence

Kutaiba Saleh - kutaiba.saleh@med.uni-jena.de.

Development of Observation Support System Using Integrated Nursing Practice Data

Ryoma Seto^a, Toshitaka Inoue^b, Nobuyuki Koeda^c,
Sayuri Yamashita^d, Kotaro Komino^d, Takahiro Yasoshima^e

^a Division of Healthcare Informatics, Faculty of Healthcare, Tokyo Healthcare University, Tokyo, Japan

^b Department of Social Welfare Science, Faculty of Health and Social Welfare, Nishikyushu University, Saga, Japan

^c Secretariat, Yao Municipal Hospital, Osaka, Japan

^d Department of Nursing, Municipal Fussa Hospital, Tokyo, Japan

^e Department of Primary Care, Sapporo Dohto Hospital, Hokkaido, Japan

Abstract

The purpose of this study was to develop a prototype nursing observation support system using integrated nursing practice data with nursing records, prescription data, and nurse call logs. These data show that the present observation system has improved. The system has the potential to provide improved observations of chest symptoms and pain management.

Keywords:

Hospital Information Systems, Nursing Record, Patient Safety

Introduction

Inpatient care support systems are rapidly getting integrated with patient care devices (PCDs) [e.g., nurse call system, fall prevention sensor, electrocardiogram (ECG) monitor]. However, these integrated systems function mainly to alert nurses and are not used to predict patient conditions. Observation is one of the most important roles in nursing care. Therefore, a “structured observation function with PCDs” is required for future inpatient care systems. The purpose of this study was to develop a prototype nursing observation support system using integrated nursing practice data with PCDs.

Methods

Data were collected at a private “A hospital (105 beds)” with internal medicine and surgical units. Nursing progress notes from May to November 2015 were analyzed using the text-mining method and converted into qualitative data. The notes were integrated with the prescription order log and the nurse call log. The nurses’ work flow was visualized via patient topics (e.g., pain, chest symptoms). The work flow from those data was discussed with an inter-professional team, and a prototype prediction algorithm was developed. This study was approved by the managing board under A hospital’s protocol.

Results

Observation of Chest Symptoms

Among 237,202 nursing records from May to November 2015, 1,942 records were related to ECG. Of the ECG-related records, 582 were normal sinus rhythm (NSR), but only 38 (6.5%) were written as “chest symptom.”

Observation of Pain

Among 237,202 nursing records from May to November 2015, 1,990 were related to “nurse call.” Text-mining picked 4,546 key terms (noun) as top 10 nouns from all “nurse call” records. Most nouns were related to humans or materials (e.g., “nurse” and “bed”); the only noun relating to symptoms was “pain.” From May to November 2015, “nurse call” was related to “pain” in 457 patients (49 patients in November). In 49 patients, the prescription data and nurse call log were connected to their nursing records. Consequently, only 12 patients (33.3%) used pain-relief medication (e.g., non-steroidal anti-inflammatory drugs). However, two stages of pain were written in their records: able or unable to bear.

Development of Observation Recommendation

The nursing practice data showed nursing observation to be improved. The inter-professional team discussed the algorithm of nursing intervention. If a nurse records information about ECG, the prototype system provides recommendations for observation, including chest symptoms. If the nurse call rings often and the nursing record includes pain, the system asks the nurse whether the patient feels pain.

Discussion

PCDs have a huge potential to increase opportunities of nursing observation. Electronic medical records also increase the quality of nursing care. However, the results of this study indicate that the nurse cannot use these patient data at full value. Therefore, the “cognitive limit” of nursing observation should be considered in the development of inpatient care systems using PCDs.

Conclusions

An observation support system using integrated nursing practice data can improve nursing observation, and the system is useful for patient safety and improvement of the quality of nursing care such as pain control.

Acknowledgements

This work was supported by a JSPS Kakenhi Grant Number 16K12222.

Address for correspondence

Prof. Ryoma Seto (rseto-ind@umin.ac.jp)

Accuracy of Decision Support Systems for Breast Cancer – Initial Results

Priscyla Waleska Simões^{a,b,c,f}, Ronaldo Borges Vicente^{a,c}, Patrícia Duarte Simões Pires^{a,c}, Maria Marlene de Souza Pires^d, Eros Comunello^e, Felipe Borges Tomaz^{a,c}, Luciane Bisognin Ceretta^b, Deborah Ribeiro de Carvalho^a

^a Grupo de Pesquisa em Tecnologia da Informação e Comunicação na Saúde, Universidade do Extremo Sul Catarinense, Brazil

^b Programa de Pós-Graduação em Saúde Coletiva (Mestrado Profissional), Universidade do Extremo Sul Catarinense, Brazil

^c Curso de Medicina, Universidade do Extremo Sul Catarinense, Brazil

^d Programa de Pós-Graduação em Ciências Médicas, Universidade Federal de Santa Catarina, Brazil

^e Vision Lab, Programa de Pós-Graduação em Computação Aplicada, Universidade do Vale do Itajaí, Brazil

^f Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas, Universidade Federal do ABC, Brazil.

Abstract

The aim of this study was to evaluate the accuracy of decision support systems in the diagnosis of breast cancer by means of a systematic review and meta-analysis of studies of diagnostic accuracy. The first step of the research, which consisted of the initial research of abstracts and titles identified from the research strategy in the databases was performed by two researchers independently. In this stage, 622 references were retrieved in the databases and, through a consensus meeting, 183 articles were selected for full reading.

Keywords:

Decision Support Systems, Clinical; Review; Meta-Analysis

Introduction

Historically to the present day, the reason for a technology not being used in clinical practice may be the lack of a robust evaluation, high cost, or because it presents itself as a very invasive method. Considering the importance of the issue in the global context, some countries in the Americas, Europe and Asia have turned their attention to the evaluation of technologies applied to the health context and implemented government agencies that specifically focus on this issue [1].

On the other hand, due to population aging in the next 20 years, the estimate for cancer cases generally suggests a 75% increase in prevalence [2]. Considering this information and in addition to the fact that breast cancer is very common among women, the concern is that there will be growth of breast cancer cases in Brazil and in the world. According to the World Health Organization (WHO) prevention strategies can not extinguish most of the cases that occur in low- and middle-income countries, where the diagnosis of the disease is made in advanced stages, so early diagnosis is necessary. In order to improve the prognosis [2].

This paper presents the initial results of the accuracy of decision support systems in the diagnosis of breast cancer by means of a systematic review and meta-analysis of studies of diagnostic accuracy.

Materials and Methods

An exhaustive search of Medline, Embase, and Gray Literature was performed for publications between 1970 through 2016, and were included diagnostic accuracy studies that evaluated individuals with breast cancer (target condition) by means of decision support systems (test in evaluation).

Results

The first step of the research, which consisted of the initial research of abstracts and titles identified from the research strategy in the databases was performed by two researchers independently. In this stage, 622 references were retrieved in the databases and, through a consensus meeting, 183 articles were selected for full reading.

Conclusion

The present project is being developed with the following steps, reading the full articles, developing the meta-analysis and evaluating the methodological quality of the included studies, which will be done through the tool Quadas 2.

Acknowledgments

Financiadora de Estudos e Projetos (FINEP) and Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina (FAPESC).

References

- [1] J.M. Stephens, B. Handke and J.A. Doshi, International survey of methods used in health technology assessment (HTA): does practice meet the principles proposed for good research. *Comp Eff Res* 2 (2012), 29-44.
- [2] Pan American Health Organization. OMS: Organização Mundial da Saúde; 2016.

Address for correspondence

Prof. Priscyla Waleska Simões, PhD - pritisimoes@gmail.com

Accessibility of Mobile Devices for Visually Impaired Users: An Evaluation of the Screen-Reader VoiceOver

Berglind Smaradottir^a, Jarle Håland^{b,c}, Santiago Martinez^b

^a Department of Information and Communication Technology, University of Agder, Grimstad, Norway

^b Department of Health and Nursing Science, University of Agder, Grimstad, Norway

^c Kongsgård School Centre, Kristiansand, Norway

Abstract

A mobile device's touchscreen allows users to use a choreography of hand gestures to interact with the user interface. A screen reader on a mobile device is designed to support the interaction of visually disabled users while using gestures. This paper presents an evaluation of VoiceOver, a screen reader in Apple Inc. products. The evaluation was a part of the research project "Visually impaired users touching the screen - a user evaluation of assistive technology".

Keywords:

Visually Impaired Persons, Medical Informatics

Introduction

Mobile technologies with touchscreen allow users to interact with the interface through hand gestures. Touchscreens may represent a significant accessibility and communication barrier to users with visual disabilities since this type of interaction does not usually provide audio or tactile feedback by default. However, there currently are in the market solutions that enable visually impaired users to use technology using voice commands. An example of this technology is the VoiceOver by Apple Inc., developed for iOS devices such as iPhone and iPad. VoiceOver synchronizes speech feedback with user hand gestures. This work presents a user evaluation of the screen reader VoiceOver, with the aim of exploring the challenges that visually disabled users experience when they interact with a touchscreen using hand gestures.

Methods

A usability evaluation in the Usability Laboratory [1] at the Centre for eHealth and Healthcare Technology at the University of Agder, Norway, was made with 6 visually impaired participants. They were asked to use a series of gestures to solve tasks related to the use of a smartphone (i.e., iPhone). Quantitative (time and attempts per gesture) and qualitative (pre- and posttest semi-structured interviews) measures were used for the analysis of the data collection.

Results

Participants reported that most of the hand gestures were generally easy to perform. However, several gestures were described as "challenging" or requiring user-training in advance. The system responded mainly correctly to the set of

gestures, but a few times the system did not respond to the hand gestures performed on the user interface (UI). For instance, several participants had problems with ending a phone call. In addition, the speech feedback responded appropriately to the gestures. However, one particular gesture (swipe gesture) was inconsistently interpreted across the tests. The speech feedback provided sufficient information to enable participants to navigate through the UI. Only a few times there was a lack of information in the speech feedback or expressions that were not understood by participants.

Conclusion

Mobile device touch screens are widely used by heterogeneous user groups, which emphasizes the importance of key human-computer interaction concepts such as accessibility and usability. Accessibility and usability evaluations are essential to improve the interface design and the interactions between mobile devices and users with physical, cognitive and/or sensory limitations. This work has summarized an analysis of the usability of the gesture-based speech-assisted interface navigation system, Apple Inc. VoiceOver. The findings showed that, in general, the system feedback was appropriate, but there were some specific functions and speech expressions that could be improved to increase effectiveness, efficiency and user satisfaction of visually impaired users.

Acknowledgements

The authors thank the visually disabled people for their disinterested participation in this study.

References

- [1] B. Smaradottir, J. Håland, S. Martinez, Å.R. Somdal and R. Fensli, Recommendations on a test infrastructure for evaluation of touchscreen assistive technology for visually impaired users, *Proceedings of the 13th Scandinavian Conference on Health Informatics*, Tromsø, Norway, 15-17 June 2015; 41-46, ISBN: 978-91-7685-985-8.

Address for correspondence

Dr. Berglind Smaradottir, Department of Information and Communication Technology, Faculty of Engineering and Science, University of Agder, PB 422, N-4604 Kristiansand, Norway.
Email: berglind.smaradottir@uia.no

Advancing Towards a CDA-Based Trauma Registry Data Submission

Martin Staemmler^a, Alexander Nischwitz^a, Markus Blätzing^b, Johannes Sturm^b

^a Medical Informatics, University of Applied Sciences, Stralsund, Germany,

^b TraumaRegistry®, Academy of the German Society for Trauma Surgery, Munich, Germany

Abstract

This poster presents an assessment to which extent the data submission to one of the largest trauma registries worldwide can be structured compliant to the clinical document architecture (CDA) and semantically annotated. Overall, complete annotation was achieved for 75% of the items, for the remaining ones annotation failed due to missing codes or concepts for individual items or values of value sets.

Keywords:

Registries, Semantics, Electronic Health Records

Introduction

Data submission to trauma registries relies either on web-based forms or files holding the items entered using a standalone application or an application integrated with a clinical information system (HIS, EMR) [1, 2]. Usually the data representation is proprietary, registry dependent, lacks semantic annotation and/or conformance to established standards regarding data types, structure (e. g. entries, sections) or building blocks (e. g. templates). The objective of this paper is to assess to which extent the data submission to one of the largest trauma registries worldwide, the German TraumaRegistry®, can be represented compliant to the above requirements.

Methods

The analysis of the trauma registry data submission revealed five sections reflecting the phases: trauma and demographic data, pre-hospital, emergency room/operating theatre, intensive care unit, diagnosis and outcome. Some of these phases include subdivisions e.g. laboratory, diagnostic procedures, emergency interventions, anticoagulation. The data types of the ca. 100 items comprise of free text, single and multiple values, scores and structured items with defined value sets. To assure maintainability the open-source tool ART-DECOR® was used. It supports the specification of data sets by providing access to HL7 templates, value sets, data types as well as to classifications and nomenclatures. The structure of the submission data set was mapped to a document template with sections for the phases and the subdivisions. For each item a data type was selected using the HL7 CDA set of data types. Furthermore, each item was semantically annotated based on the concept of coded values. This annotation has also been extended to individual values of value sets. Project related object identifiers were assigned to items and values sets provided that there was no existing assignment. Overall, a table was used to monitor the status of each item regarding semantic annotation for later assessment.

Results

The phases and subdivisions were reflected in CDA sections. Items led to entries. Semantic annotation was achieved using ICD-10, LOINC and SNOMED CT (only precoordination). Overall, 75% of all items could be annotated completely, exhibiting the highest rates with the pre-clinical and the emergency room/operating theatre phases. Reasons for annotation failure were missing or inapt codes for items (21%) or values within a value set (4%). For example, the NACA score is missing in SNOMED or LOINC, values of the value set “type of accident” could not fully be mapped to existing SNOMED concepts. ART-DECOR® was versatile for structuring and semantic annotation, its specification file (decor.xml) serves for later verification and instantiation of a CDA document.

Discussion and Conclusions

In 2016 HL7 published a CDA implementation guide for trauma registry data submission for the NTDS [3], taking a similar approach, but based on a partially different sets of items. The more extensive Canadian approach [4] uses proprietary data types but lacks semantic annotation. Other, mainly IHE profiles e.g. XDS-MS, EDR in the PCC domain [5] rely on CDA but are focused on the emergency department. As a result, the two CDA based approaches better comply to the registry requirements while taking the different realms into account.

References

- [1] Academy for Trauma Surgery, 20 years TraumaRegister DGU: Development, aims and structure, Injury, Int. J. Care Injured **45S** (2014) S6–S13.
- [2] American College of Surgeons, NTDS Data Dictionary, 2017 Admissions, www.ntdsdictionary.org/dataelements/documents/NTDSDataDictionary-2017Admissions.pdf
- [3] HL7 International, HL7 CDA® R2 Implementation Guide: Trauma Registry Data Submission, Release 1, US Realm, August 2016, www.hl7.org
- [4] Canadian Institute for Health Information, Ontario Trauma Registry Comprehensive Data Set Data Dictionary, May 2014, www.cihi.ca/en/services_otr_cds_dict_en.pdf
- [5] IHE International, TF-PCC, Vol 1, Profiles: Cross-Enterprise Sharing of Medical Summaries, Emergency Department Referral, Nov. 2016, www.ihe.net

Address for correspondence

Prof. Dr. Martin Staemmler, martin.staemmler@fh-stralsund.de

Patient Accessibility to Hospitals in Winter Road Conditions: GIS-Based Analysis Using Car Navigation Probe Data

Takumi Tanikawa^a, Hisateru Ohba^a, Ayako Yagahara^a, Katsuhiko Ogasawara^b

^aFaculty of Health Sciences, Hokkaido University of Science, Sapporo, Hokkaido, Japan,

^bGraduate School of Health Sciences, Hokkaido University, Sapporo, Hokkaido, Japan

Abstract

The purpose of this study was to estimate geographical patient flow to hospitals during winter seasons using simulation analysis. We used probe data collected from car navigation systems and performed a Geographical Information System (GIS)-based analysis to determine the relationship between travel time to hospitals and winter road conditions. Accessibility to hospitals based on travel time in summer and winter was overlayed on a map to demonstrate the increase in travel time during winter.

Keywords:

Health Services Research, Transportation

Introduction

In snowy areas, road surface conditions can significantly deteriorate in the winter. The decline in traffic flow due to snowfall and freezing of roadways leads to a decreased accessibility to hospitals, and increase in travel time can be very problematic during an emergency. In this study, we analyzed the difference in travel time to hospitals based on weather and road conditions in summer and winter.

Methods

We examined 21 secondary medical service areas (geographical administrative unit for health policy) of Hokkaido Prefecture, Japan. It was assumed that outpatients traveled for their hospital visit by car. Several weather scenarios were defined to determine cost factors. The area examined was also classified into urban and suburban areas to set a cost value. Travel time in winter equalled that in summer multiplied by the cost value. Based on the scenario, the cost was set to 1 for roads in summer and from 1 to 3 for roads in winter. Additionally, actual cost values were determined using probe data of a week in each of summer and winter collected from a car navigation system.

We used the geographic information system Arc GIS 10.1 and road network data from ArcGIS data collection 2013 (Esri). The 15-min to 120-min travel time areas were projected on a map, and the population coverage ratio in each area was calculated by each secondary medical service area.

Results

In winter, throughout Hokkaido, the cost value calculated using probe data was 1.15 in urban areas and 1.29 in suburban areas. The population coverage ratio for the 60-min travel time area was decreased from 90.4% in summer to 87.9% in winter

in the Sapporo medical service area, while for the 30-min area, it was decreased from 71.5% in summer to 64.8% in winter.

Discussion

Data used for formulating medical plans is usually derived from investigations during the summer season. However, weather and road surface conditions in winter may restrain patient visiting behavior; thus, winter conditions in snowy areas should be considered while formulating medical plans. The results of this study provide a visual demonstration based on the weather scenario and the probe data that patient travel time to core hospitals becomes longer in winter than in summer. A limitation of this study is that only two weeks probe data was taken, thus the cost value may change depending on collected periods.

Conclusions

In this study, we used GIS data to create a map displaying the increase in travel time to core hospitals in Hokkaido in winter. We also estimate the time required to travel to core hospitals based on weather and road conditions in winter as well as the population coverage ratio for the secondary medical service area based on weather scenario and car travel data.

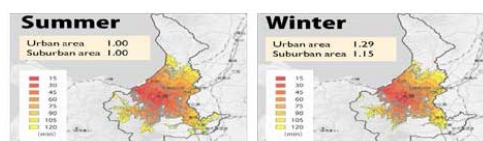


Figure 1 – 15(red)–120(yellow)-min travel area (Sapporo secondary medical service area)

Address for correspondence

Takumi Tanikawa, Ph.D., E-mail: tanikawa-t@hus.ac.jp

Implementation of an Outsourced Transnational Service of Clinical Decision Support System

Vania Teixeira^a, Luciana Rubin^b, Romina Rebrij^b, Alvaro Tamborindeguy^a,
Rodrigo Martinez^a, Juan Carlos Bacigalupo^a, Daniel Luna^b

^aUniversal Soluciones Tecnológicas, Uruguay,

^bDepartment of Biomedical Informatics, Hospital Italiano de Buenos Aires, Argentina

Abstract

The aim of this study is to describe the implementation and evaluation of an outsourced Clinical Decision Support Systems (CDSS) service of drug-drug interaction (DDI) alerts in an Uruguayan outpatient healthcare network. A cross-sectional study was developed. 1.5 alerts were triggered of every 1000 prescriptions. Clinicians accepted 44 % of the total alerts. In conclusion, the implementation of CDSS was achievable.

Keywords:

Decision support systems, Clinical; Drug Interactions; Outsourced Services

Introduction

DDI are causes of preventable adverse drug events (ADEs). DDI type D can produce mayor ADEs and type X may be mortal. DDI are related to 2%-3% of the hospital admissions [1]. CDSS can reduce ADEs caused by DDI, by alerting the physician about them [2]. Universal Soluciones Tecnológicas (UST) is a company that provides IT to an Uruguayan private healthcare network with 300 thousand consumers. Systematization of medical terminology has been introduced in the Electronic Medical Records (EMR), consuming Hospital Italiano de Buenos Aires (HIBA) terminology web services. The purpose of this study was to implement and to evaluate an outsourced CDSS service of DDI alert in an outpatient healthcare network.

Methods

A cross-sectional study was carried out taking 5 months before and 5 months after the implementation of the IDD alerts, in the health network of UST in Uruguay. The EMR was audited in case of overridden DDI alerts to evaluate user's override's motives. The prevalence of DDI was compared before and after CDSS implementation. Data analysis was done with EPIDAT version 3.1 with 95% confidence level.

Results

CDSS services including DDI alerts of type D and X was developed by HIBA and uploaded to a web service. UST started consuming from a development environment and following HIBA design guides, for the graphic alert's interface. Two months later with the satisfactory tests, the CDSS services went live. EMR users (physicians) received capacitation through the application itself. During the 5 months after CDSS implementation 1.5 DDI alerts were

triggered every 1000 prescription. Physicians cancelled 44% of prescriptions with DDI alert and override 54%. Override justifications were: 42% "other" explained by free text, 32,5% "limited time", 13% "benefits/risks", 7% "tolerance" and 5,5% "no alternatives available". Compared to prescriptions performed before the CDSS implementation, DDI after CDSS alerts implementation were approximately 80 % lower (RR: 0.21; CI_{95%}: 0.18-0.26; P<0.0001). The prevalence of patients with DDI over the total of patients with any drug prescription was 0.50% before CDSS alerts compared to 0.30 % with CDSS alerts (RR: 0.68; CI_{95%}: 0.57-0.82; P<0.0001).

Discussion

The outsourced CDSS service was effectively and efficiently implemented. The overridden alerts were mainly attributable to a gap between knowledge and practice. Further investigations are necessary to improve the acceptance. A decrease in the DDI prevalence was associated to the implementation of CDSS alerts. However, analytical studies need to be conducted to investigate the potential causality. CDSS showed to be a tool for learning and updating knowledge for a variety of users across national and institutional boundaries. Users trained virtually had heterogeneous knowledge and capabilities and resulted in a minor limitation in the implementation.

Conclusion

The implementation of the outsourced CDSS was feasible, without major technical or professional difficulties.

Acknowledgments

Authors thank Sonia Benitez and Daniel Rizzato for their support in this writing process.

References

- [1] P.J. McDonnell, and M.R. Jacobs, Hospital Admissions Resulting from Preventable Adverse Drug Reactions, *Ann. Pharmacother.* 36 (2002) 1331–1336.
- [2] AHRQ, Clinical Decision Support (CDS), (2016). <https://healthit.ahrq.gov/ahrq-funded-projects/clinical-decision-support-cds> (accessed April 11, 2017).

Address for correspondence

vania.teixeira@universal.com.uy

Methods for Detecting Malfunctions in Clinical Decision Support Systems

Adam Wright, PhD^{a,b,c}, Trang T. Hickman, MPH^{a,c}, Dustin McEvoy^{a,c}, Skye Aaron, Angela Ai^{a,c}, Joan S. Ash^e, Jan Marie Andersen^{a,c}, Rachel Ramoni, DMD, ScD^{c,d}, Milos Hauskrecht, PhD^f, Peter Embi, MD^g, Richard Schreiber, MD^h, Dean F. Sittig, PhDⁱ, David W. Bates, MD, MSc^{a,b,c}

^a Brigham and Women's Hospital, Boston, MA ^b Harvard Medical School, Boston, MA ^c Partners HealthCare, Boston, MA ^d Harvard School of Dental Medicine, Boston, MA ^e Oregon Health and Science University, Portland, OR ^f University of Pittsburgh, Pittsburgh, PA ^g Ohio State University, Columbus, OH ^h Holy Spirit Hospital, Danville, PA, ⁱ University of Texas Health Sci. Center, Houston, TX

Abstract:

Clinical decision support systems, when used effectively, can improve the quality of care. However, such systems can malfunction, and these malfunctions can be difficult to detect. In this poster, we describe four methods of detecting and resolving issues with clinical decision support: 1) statistical anomaly detection, 2) visual analytics and dashboards, 3) user feedback analysis, 4) taxonomization of failure modes/effects.

Keywords:

Expert Systems, Electronic Health Records, Safety Management

Introduction

Significant evidence suggests that clinical decision support (CDS) systems, when used effectively, can improve health care quality, safety, and effectiveness [1]. However, unintended consequences and safety issues around CDS have been reported [2]. In prior qualitative studies, we identified several cases where CDS systems failed, in many cases without being detected [3]. Here, we focused on studying the identification, causes, and effects of CDS failures.

Methods

CDS systems issues were identified using many approaches:

1. Statistical anomaly detection on alert firing logs
2. Visual analytics/dashboards for knowledge engineers
3. Free-text analysis of user feedback to identify problematic logic
4. Taxonomization of failure modes and effects to help us develop prevention strategies

These methods have been used at six healthcare systems, including a mix of academic medical centers, integrated delivery systems, and community health care systems.

Results

Statistical anomaly detection: A mix of simple heuristic anomaly detection, Poisson change-point analysis, and Seasonal Trend decomposition using Loess identified potential issues from alert firing logs. Issues included errors due to changes in underlying codes, system upgrades, and service malfunctions.

Visual analytics and dashboards: Several dashboards for visualizing CDS alert firing and acceptance rates using D3.js and Google Charts were developed. Detected issues included mis-configured alerts firing too often, and system upgrade errors.

User feedback analysis: Rules and sentiment analysis were used to identify user free-text comments from override reasons in response to CDS alerts suggesting anger or frustration. These methods have identified failures related to order detection, medication classification, and clinical logic errors.

Taxonomization of failure modes and effects: From analyzing 70 known CDS malfunctions, we developed a tri-axial taxonomy of causes, modes of discovery, and effects. This taxonomy can identify best practices for preventing and detecting CDS malfunctions.

Conclusions

A multi-modal approach is necessary for preventing, detecting, and studying CDS malfunctions. Our methods detected a range of issues at six sites as well as identified malfunctions other methods missed. We are now identifying best practices for avoiding CDS malfunctions, including system testing, log file monitoring, CDS governance, organizational policies and procedures, developer and user training, and new analysis and visualization tools.

Acknowledgements

Research was supported by the National Library of Medicine of the National Institutes of Health (NIH) under Award Number R01LM011966. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

- [1] K. Kawamoto, C.A. Houlihan, E.A. Balas EA, D.F. Lobach, Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* **330** (2005), 765.
- [2] J.S. Ash, D.F. Sittig, E.M. Campbell, K.P. Guappone, R.H. Dykstra, Some unintended consequences of clinical decision support systems. *AMIA Annual Symp Proc* **2007** (2007), 26-30.
- [3] A. Wright, T.T. Hickman, D. McEvoy D, S.S. Aaron, A. Ai, J.M. Andersen, S. Hussain, R. Ramoni, J. Fiskio, D.W. Bates, Analysis of clinical decision support system malfunctions: a case series and survey. *J Am Med Inform Assoc* **23** (2016), 1068-1076.

Address for correspondence

Adam Wright, PhD, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115, USA; awright@bwh.harvard.edu

Data Mining on Numeric Error in Computerized Physician Order Entry System Prescriptions

Wu Xue^{a,b}, Wu Changxu^c

^a Peking University, School of Nursing, Beijing, China,

^b Institute of Psychology, Chinese Academy of Sciences, Beijing, China

^c Department of Industrial and Systems Engineering, State University of New York (SUNY)-Buffalo, NY, USA

Abstract

This study revealed the numeric error patterns related to dosage when doctors prescribed in computerized physician order entry system. Error categories showed that the '6', '7', and '9' key produced a higher incidence of errors in Numpad typing, while the '2', '3', and '0' key produced a higher incidence of errors in main keyboard digit line typing. Errors categorized as omission and substitution were higher in prevalence than transposition and intrusion.

Keywords

Medical order entry systems; Medication errors; Numeric

Introduction

Computerized physician order entry (CPOE) system can significantly reduce certain types of prescription errors. However, prescription errors still occur. Various factors such as the numeric input methods in human computer interaction (HCI) produce different error rates and types. This study aimed to categorize the types of errors on different numeric keys when doctors prescribed using CPOE. Several error patterns in this study were compared with Salthouse's findings [1] and Lin & Wu's research [2].

Methods

Data mining techniques, including clustering and classification, were used to explore the incidence of four error categories. All the errors were categorized into four types[1]: 1. (Omission), e.g. the target number was 50, while the entry number was 5, recorded '0' key occurred 1 Omis error; 2. (Substitution), e.g. the target number was 50, while the entry number was 55, recorded '0' key occurred 1 Subs error; 3. (Transposition), e.g. the target number was 25, while the entry number was 52, recorded '2' key occurred 1 Trans error; 4. (Intrusion), e.g. the target number was 50, while the entry number was 500, recorded '0' key occurred 1 Intru error. The error rate of 0-9 key in Numpad and main keyboard digit line was calculated respectively.

Results

A total of 41,100 dosages were collected from a simulated CPOE system, of which 954 sets of numbers involving 1,737 numeric errors were identified. Measuring the composition of errors, omission and substitution were higher than transposition and intrusion (Figure 1). The '6', '7', and '9' key had a higher incidence of errors using the Numpad, while the

'2', '3', and '0' keys had a higher incidence of errors in main keyboard digit line typing (Figures 2 & 3).

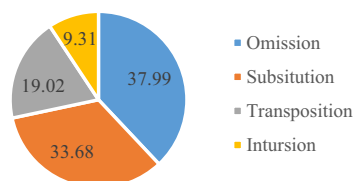


Figure 1– Composition of errors

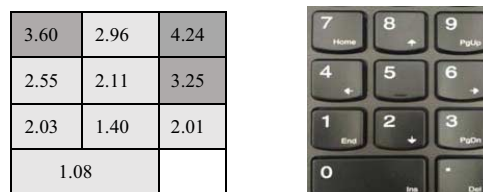


Figure 2– Error rate of different keys with Numpad

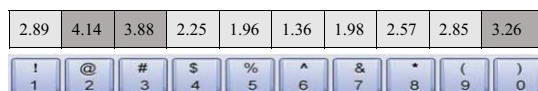


Figure 3– Error rate of different keys with main keyboard digit line

Conclusions

The omission and substitution error rates were higher than transposition and intrusion. Significant difference existed in error rates of each key between keyer layouts.

References

- [1] T.A. Salthouse, Perceptual cognitive and motoric aspects of transcription typing, *Psychol. Bull.* **99** (1986), 303-319.
- [2] C.J. Lin & C. Wu, Factors affecting numerical typing performance of young adults in a hear-and-type task, *Ergonomics* **54** (2011), 1159-1174.

Address for correspondence

Wu Changxu: Department of Industrial and Systems Engineering, State University of New York (SUNY)-Buffalo, 319 Bell Hall, University at Buffalo, NY14228, USA. E-mail: changxu.wu@gmail.com, Tel: 716-645-4715

Development of Multiple Wireless Network Environments for Hospitals

Yoshinori Yamashita^a, Tatoku Ogaito^a

^a Department of Medical Informatics, University of Fukui Hospital, FUKUI, Japan

Abstract

The use of smart devices in medical information systems is progressing. In addition, correspondence of communication functions in medical equipment is advancing. For this reason, and in order to establish the ubiquitous environment in hospitals, we developed an environment in which multiple wireless communications can be used.

Keywords:

Wireless Technology; Risk Management; Management Information Systems.

Introduction

In our hospital, the private cloud system has changed with virtualization of the hospital information system [1]. As a result, the miniaturization of terminals and the use of smart devices has expanded. In addition, data acquisition from medical devices by wireless communication has been accumulating. In particular, we have developed a communication infrastructure to utilize communication with smart devices and medical devices. However, in communications from medical devices and communication from event monitoring devices, it is necessary to reduce electric power in communication, so the use of power saving communication is required. For this reason, we can no longer use the Wifi communication that we have used so far and must switch to a system like IEEE 802.15. With this system we are able to collect data, alarm events, and identify location information [2].

Methods

We have made it possible to use the following as multiple wireless communication in the hospital:

1. Using Wireless LAN (Wifi)
2. Communication by IEEE 802.15
(IEEE 802.15.1 Bluetooth & IEEE 802.15.4 ZigBee)
3. Function for Indoor GPS(IMES)

We developed a gateway device called "IoT gateway" that relays these communications to each other (Figure 1). By using this gateway device, we collect data by relaying information from medical equipment and sensors. Also, it is possible to perform positioning with BLE beacon or GPS radio waves for smart devices.

By installing this gateway device at 1000 places throughout the hospital (i.e. in each room and hallway in the hospital), we can obtain the position with the collected data.

Results

In previous methods, communication methods other than Wifi were operated on a point-to-point basis, but with this it was

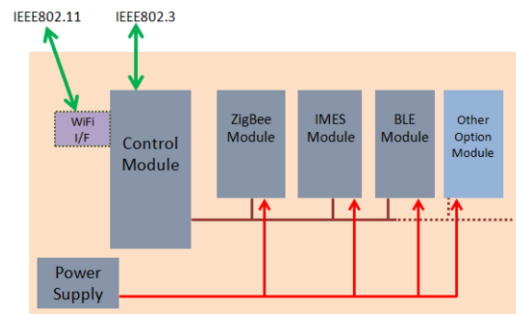


Figure 1 – Scheme of Gateway Device

possible to establish as an infrastructure that could be used all the time. We have been able to handle data by wireless communication anywhere in the hospital. Therefore, it is possible to utilize a medical device with a communication function. In addition, because it has the function of providing location information, it is now possible to specify the location where data is collected. As another use, it was also possible to restrict data access by location. As a result of the development of this gateway device, wireless communication for embedded devices can be used steadily and expansion of wireless compatible devices in the future can also be done. Furthermore, it is possible to collect data with location information.

Conclusion

With this method, location information can be obtained together with collected data, so that data analysis can be performed along with location. Therefore, it can be used for analyzing accidents and workflow information. By using the results from this study we were able to change the work of nurses and succeeded in reducing overtime.

References

- [1] Y. Yamashita and T. Ogaito, Rebuilding and the private cloud of the hospital information system by the virtualization technology, In: *MEDINFO2013 Proceedings* (2013):1183-1183.
- [2] Y. Yamashita, T. Ogaito and S. Kasamatsu, Development of medical equipment alarm monitoring system, In: *MEDINFO2013 Proceedings* (2013):1195-1195.

Address for correspondence

Yoshinori Yamashita : e-mail: yyama@u-fukui.ac.jp

Core Curriculum Comparisons in Master Degree Programs for Patient Safety

QiYan Zhang, Hiroshi Takeda

Graduate School of Health Sciences, Jikei Institute, Osaka, Japan

Abstract

A content analysis was performed to investigate core curriculum in master degree programs for patient safety. Twenty programs from six countries and 179 core curriculum/modules were extracted and reviewed. The curricula were compared to the topics recommended in WHO's Multi-professional Patient Safety Curriculum Guide and the core content relevant to patient safety issues in IMIA/AMIA recommendations for health informatics education. The differences between the identified curricula with those addressed in the aforementioned Guide/Recommendations were examined.

Keywords:

Patient Safety; Education; Curriculum

Introduction

Some master degree programs have been established to target all healthcare professionals recently. Studies indicating the ways in which “patient safety” are incorporated within such programs are still limited. So far, there is no consensus exist in this domain. WHO's Multi-professional Patient Safety Curriculum Guide[1] is considered as the gold standard. Meanwhile, IMIA and AMIA also addressed some core content relevant to patient safety issues in their recommendations on education for health informatics [2] [3]. In this study, we conducted a content analysis to investigate core curriculum developed in the master degree programs for patient safety, in general, to understand: 1) the coincidence degree of the existing curricula with the WHO's Guide; 2) the differences between these curricula with those addressed in the aforementioned Guide/Recommendations.

Methods

We searched the literature using Pubmed and Google Scholar to collect and extract the study subjects. We excluded those curricula, which were developed either in certificate programs or in residency training/fellowship programs. Then we drew up a list of the identified curriculum/modules for evaluation. Program Title, Name of the University/Institute, Country, Study Type, Durations, Degree Structure, Course Name were used to describe each program. Finally, we reviewed the contents of each curriculum and compared them with the topics recommended in WHO's Guide. Also, we examined the differences between the identified curricula with those addressed in the corresponding Guide/Recommendations.

Results

Twenty master degree programs from six countries were identified: US (8), UK (7), Canada (2), Spain (1), Japan (1) and United Arab Emirate (1). 179 core curriculum/modules were extracted and reviewed. The coincidence degrees with WHO's recommendations were: “Using quality improvement methods to improve care” 100%; “What is patient safety?” 70%; “Understanding and managing clinical risk” 65%; “Understanding systems and the effect of complexity on patient care” 50%; “Being an effective team player” 50%; “Why applying human factors is important for patient safety” 45%; “Learning from errors to prevent harm” 45%; “Engaging with patients and carers” 45%; “Improving medication safety” 20%; “Infection prevention and control” 5% and “Patient safety and invasive procedures” 5%, respectively. Health informatics topics, e.g. electronic health records, clinical informatics, decision support processes, human-computer interface design, and usability were involved in 40% of these curricula. Beyond topics covered by WHO's Guide and IMIA/AMIA recommendations, research methodologies, social issues regarding policy/leadership/legal/law/regulatory topics were typically focused in the existing curricula.

Conclusions

The identified curricula in the existing programs reflect the educational trends in the field.

References

- [1] World Health Organization. Multi-professional Patient Safety Curriculum Guide. WHO Press, 2011.
- [2] Mantas J, Ammenwerth E, Demiris G, et al; IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health Informatics. First Revision. *Methods Inf Med* 2010; 49:105e20
- [3] Gardner RM, Overhage JM, Steen EB, et al; AMIA Board of Directors. Core content for the subspecialty of clinical informatics. *J Am Med Inform Assoc* 2009; 16:153e7

Address for correspondence

Dr. Qiyan Zhang (e-mail address: qyhz.louise@gmail.com).

Subject Index

A

access to information 89, 1305, 1308
 accidental falls 828, 973, 1043, 1365
 acute coronary syndrome 398
 addictive 1272
 adherence 1306
 administration 1251
 adult day care centers 1215
 adverse drug reaction reporting systems 1268
 adverse event 1262
 Africa South of the Sahara 965
 aged 161, 1307
 ageing 569
 AIDS 1352
 algorithms 366, 531, 599, 843, 873
 Alzheimer disease 1249
 anesthesia 1256
 anonymization 1303
 anterior cruciate ligament 1217
 anticoagulants 1024
 antihypertensive agents 1200
 aphasia 1374
 application 240
 appointments and schedules 1243
 archetype 1344
 archiving 337
 Argentina 835
 arrhythmia 118
 arthritis 1223
 artificial intelligence 447, 559, 960, 1249, 1288, 1331
 assisted living facilities 166
 asthma 1170
 atrial fibrillation 1024
 attitudes 1314
 Australia 471
 autism spectrum disorder 1133
 automatic encoding 1260
 automation 1362

B

barcoding 49
 barriers 1314
 bayes theorem 1291, 1355, 1372
 BCMA 1370
 behavior 1272
 benchmarking 64, 1266
 bibliometrics 853
 big knowledge 978
 biological ontologies 925, 1153

biomarkers 98
 biomedical ontologies 457, 838, 863
 biomedical research 341, 1276
 biomedical technology 496, 878
 biometry 1053
 biomimetics 1271
 blood glucose 619, 684, 1058
 blood transfusion 994
 Brazil 1253
 breast cancer 156
 breast feeding 49
 breast neoplasms 1267
 bundled payment 1359

C

cancer 126
 cancer staging 1323
 capacity building 1245
 carcinoma 634, 1291
 care coordination 103
 case based reasoning 1371
 cell phones 1301
 censuses 1299
 cerebral infarction 1280
 certification 783
 change management 813
 chemistry 1372
 China 427
 Chinese traditional 653, 1014
 chronic 1232, 1354
 chronic disease 136, 1180, 1230
 chronic pain 275
 classification 1260, 1261
 clinical 294, 447, 700, 882, 983, 1148, 1190, 1230, 1277, 1322, 1323, 1324, 1329, 1336, 1355, 1363, 1372, 1378, 1380, 1384
 clinical coding 427, 965
 clinical decision support systems 1153, 1175, 1255, 1326
 clinical decision-making 634, 1099, 1274, 1281
 clinical laboratory information systems 1063, 1231
 clinical pathway 1344
 clinical phenotyping 432
 clinical protocols 59
 clinical research informatics 950
 clinical study 516
 clinical trial(s) 858, 1262, 1265, 1341
 cloud computing 35

cluster analysis 491, 639, 667, 1185
 co-word analysis 1293
 cognition 1324
 cognitive dysfunction 1219
 cohort studies 108
 colorectal neoplasms 1263, 1270
 commerce 773
 common data elements 337, 905, 1175
 community health services 714, 1362
 community participation 773
 comorbidity 1358
 computational biology 375
 computer games 74
 computer graphics 672, 748, 1138, 1148
 computer security 1315
 computer simulation 1242, 1298, 1321
 computer systems 1213, 1353
 computer-assisted drug therapy 1222
 computer-assisted 244, 1148, 1322, 1369, 1374
 computer-assisted decision making 332
 computer-assisted instruction 1304
 computerized 1258
 computerized medical records systems 1340
 computers 212, 589, 1339, 1376
 computers/utilization 79
 computing methodologies 166, 1099
 cone-beam computed tomography 1029, 1250
 confidentiality 1303
 conjugate vaccines 40
 consumer health informatics 126
 consumer health information 89, 178, 216
 continuity of patient care 696
 contraception 1287
 controlled 235, 511, 843, 887, 1327, 1332, 1334
 cost effectiveness 1103
 cost of illness 1358
 critical care 1319
 critical pathways 649, 672
 cross infection 743
 cross-sectional studies 122

crowdsourcing 511, 905
 curriculum 1388
 curriculum (D003479) 728

D

data accuracy 471, 573, 1300
 data annotation 432
 data anonymization 393, 704
 data collection 1143, 1309
 data display 308, 422, 578, 684
 data element 1344
 data encryption 1133
 data interpretation 662, 868
 data mining 308, 322, 346, 403, 412, 531, 544, 624, 649, 653, 873, 1075, 1267, 1268, 1282, 1287, 1288, 1361
 data quality 1352
 data systems 516
 database management system 554
 database(s) 462, 467, 481, 1275, 1289, 1338
 decision making 69, 84, 803, 1148, 1233, 1322, 1324
 decision support 40, 1063
 decision support system(s) 294, 447, 700, 882, 983, 1148, 1180, 1190, 1230, 1277, 1322, 1323, 1324, 1329, 1336, 1355, 1363, 1367, 1378, 1380, 1384
 decision support techniques 54, 1283, 1322
 deglutition disorders 1195
 delivery of health care 69, 467, 586, 1229, 1310, 1343
 dementia 569, 788
 demography 536
 dental implants 1166
 dental informatics 1290
 dental records 1166, 1281
 depression 1224
 depressive disorder 249, 1272
 developing countries 718, 738
 diabetes 379, 1209, 1210
 diabetes mellitus 188, 263, 366, 619, 684, 1211
 diabetic retinopathy 289, 559
 diagnosis 356, 384
 diagnostic errors 1364
 diagnostic imaging 244
 dictionaries as topic 940
 dietary supplements 614
 digital 1325
 digital divide 1301, 1310
 digital terrestrial
 broadcasting 1366
 disasters 1214
 disease management 915, 1014
 disease outbreak(s) 408, 619
 dissertation 1293

distance 753
 documentation 1128, 1170, 1237
 drug interactions 748, 960, 1085, 1384
 drug monitoring 1123
 drug therapy 1369
 drug utilization 1259
 drug-related side effects and
 adverse reactions 531, 549, 1123

E

ecosystem 188
 education 161, 679, 753, 1245, 1256, 1320, 1388
 education of patients 1180
 educational status 793
 efficiency 1212
 elderly 136
 electric data caputring 1279
 electric stimulation therapy 25
 electrocardiography 118, 526
 electromyographic sensor 1225
 electronic data capture 1254
 electronic dental record 1290
 electronic health record(s) (EHR)
 35, 45, 64, 141, 170, 174, 235, 308, 341, 370, 393, 432, 462, 486, 516, 539, 544, 629, 689, 709, 723, 768, 778, 798, 808, 813, 935, 945, 950, 999, 1019, 1033, 1038, 1043, 1058, 1099, 1108, 1113, 1118, 1128, 1138, 1170, 1227, 1231, 1238, 1255, 1257, 1264, 1266, 1270, 1277, 1283, 1302, 1305, 1308, 1312, 1321, 1328, 1329, 1373, 1382, 1385
 electronic mail 1315
 emergencies 989
 emergency medical services 1033
 emergency medicine 1367
 emergency service 230, 1289
 emergency treatment 230, 1304
 emotional intelligence 1235
 employment (D004651) 728
 environmental exposure 457, 496
 equipment and supplies 1345
 ethics 718, 823
 evaluation studies 298, 1269
 evaluation studies as topic 644
 evidence-based medicine 1347
 evidence-based practice 1335
 exercise therapy 1226
 expert systems 882, 1009, 1085, 1123, 1185, 1355, 1385

F

factual 1275, 1289
 fall detection 1225

financial support 714
 fitness tracker(s) 108, 146
 flowchart 59
 follow-up studies 1090
 formative feedback 447
 frail elderly 327

G

game telemetry analysis 183
 gene expression profiling 1267
 genetic counseling 1278
 genomics 89, 700
 geographic mapping 408, 569
 geriatric research 1265
 glaucoma 318
 goals 773
 guideline 1376

H

harm reduction 1378
 health 738
 health care 1090
 health care behavior 1311
 health care rationing 989
 health care surveys 689
 health disparities 1265
 health education 1244
 health expenditure 738
 health information exchange 230, 235, 723, 798, 1103, 1236, 1357, 1373
 health information systems 15, 45, 225, 442, 763, 835, 1053, 1066, 1143, 1300, 1331, 1377
 health level seven 263, 868
 health literacy 122, 178, 1211, 1339
 health manpower 1299
 health policy 40, 1312
 health promotion 146
 health records 202, 221, 1234
 health services 64, 258, 793
 health services accessibility 403
 health services for the aged 151
 health services needs and demand 1158, 1241
 health services research 1383
 health technology 1216
 healthcare 20, 1225
 healthcare system 813
 heart failure 506
 heart rate 98, 526
 hematologic tests 1274
 hepatitis B 1293
 hepatocellular carcinoma 667
 hereditary cancer syndrome 1278
 historiography 437
 history of nursing 1297
 HIV 1242
 home care services 1241

hospital information systems 778
 hospital(s) 230, 1231, 1289
 hospital administration 84
 hospital departments 1286
 hospital information systems 554,
 1058, 1279, 1313, 1360, 1368,
 1379
 hospital mortality 476
 human activity 108
 human phenotype ontology 581,
 1285
 humanism 768
 humans 151
 hypertension 1200

I

image processing 244
 imaging 1237
 incidental findings 828
 infection control 1009, 1190, 1377
 informatics 113, 212, 341, 679,
 905, 1118, 1306, 1351, 1356,
 1375
 information dissemination 303,
 462, 1242
 information extraction 521
 information literacy 1224
 information management 156, 170,
 1238, 1338
 information needs 126
 information retrieval 351
 information services 1310
 information storage and retrieval
 207, 303, 388, 422, 521, 549,
 629, 896, 1043, 1048, 1269,
 1275, 1276, 1283
 information system(s) 313, 714,
 718, 803, 1103, 1243, 1259,
 1311, 1352
 information systems success
 model 1370
 infrared rays 384
 instant messaging (IM) 818
 insurance 738
 integrated health care
 systems 1228
 intensive care unit 1271, 1319,
 1346
 international classification of
 diseases 427
 international normalized ratio 1222
 Internet 1004, 1212, 1224, 1301
 intersectoral collaboration 788,
 1337
 interstitial cystitis 94
 intravenous 1251

K

knowledge base(s) 733, 896, 940,
 1009

knowledge management 803, 838,
 878
 knowledge managementsep
 information storage and
 retrieval 983

L

laboratory critical values 1190
 learning 1284
 length of stay 1019
 libraries 1325
 life style 146
 linguistics 346
 logical observation identifiers
 names and codes 808, 1333
 long-term care 1215
 longitudinal studies 10
 lymph nodes 634

M

Machine Learning 113, 294, 356,
 388, 476, 491, 609, 667, 882,
 910, 1070, 1218, 1249, 1260,
 1263, 1273, 1280, 1288
 major 249
 managed care programs 578
 management information systems
 1215, 1387
 mapping 1342
 mass screening 1328
 mathematical computing 649
 meaningful use 1264
 medical 1256, 1302, 1320
 medical controlling 1289
 medical errors 15, 733, 1048,
 1163, 1345
 medical health records 1240
 medical informatics 94, 151, 178,
 258, 285, 327, 375, 388, 403,
 452, 457, 496, 511, 564, 624,
 693, 733, 753, 758, 783, 878,
 968, 1108, 1213, 1222, 1287,
 1381
 medical informatics
 (D008490) 728
 medical informatics application(s)
 207, 467, 1090, 1309
 medical informatics
 computing 452
 medical order entry systems 1118,
 1386
 medical records 1228, 1358
 medical records systems 1238,
 1258
 medical subject headings 662, 968
 medication adherence 1200
 medication administration
 safety 1370
 medication errors 892, 1095, 1386
 medication systems 1095

medicine 653, 1014
 medicine chinese 1257
 mental health services 249
 meta-analysis 1376, 1380
 metadata 225, 313, 935, 1337,
 1347
 microanuerysm 289
 MIMIC 1271
 mining 1253
 mobile applications 25, 54, 59,
 131, 188, 221, 253, 743, 1221,
 1227, 1232, 1241, 1320, 1363
 mobile apps 183
 mobile health (mHealth) 10, 136,
 1209, 1210
 mobile learning 1314
 models 84, 174, 506, 662, 696,
 778, 915, 1323
 monitoring 379
 mood disorders 280
 motivation 901
 motor skills 74
 multidetector computed
 tomography 1029
 multimedia 1220
 multiple chronic conditions 672
 music therapy 54
 myocardial infarction 476

N

names 965
 national health programs 1307
 natural language processing 280,
 285, 294, 298, 346, 351, 356,
 370, 393, 452, 481, 501, 521,
 539, 594, 604, 614, 629, 644,
 657, 838, 848, 910, 1235,
 1252, 1269, 1290, 1292, 1361
 neoplasms 131
 neural network(s) (computer) 318,
 559, 604, 1226, 1274, 1277
 neuroscience 5
 non-small-cell lung 1291
 numeric 1386
 nurses 212, 1340, 1353
 nursing 823, 1243, 1284, 1298,
 1351
 nursing homes 892
 nursing informatics 758, 1245,
 1297
 nursing process 1014
 nursing record(s) 1258, 1365, 1379
 nursing undergraduates 1314

O

observational study 887, 1327
 obstetrics 1019
 obstructive sleep apnea 268
 occupations 486
 ontology(ies) 910, 930, 960

ontology quality assurance 1330
 ontology summarization 978
 ontology visualization 1330
 open access publishing 901
 organization and administration
 693, 1357
 organizational affiliation 1338
 organizations 758
 osteosarcoma 1233
 outsourced services 1384

P

pain 1261
 pancreatic neoplasms 1372
 panoramic 1029, 1250
 parenteral nutrition 1038
 Parkinson disease 141
 pathology staging 1273
 patient access to records 79
 patient care team 412
 patient handoff 999, 1302
 patient harm 1075
 patient identification systems 994,
 1053
 patient medication knowledge 748
 patient monitoring 198
 patient outcome assessment 935
 patient participation 193, 1223,
 1305, 1308
 patient portals 723, 798
 patient readmission 506
 patient reported outcome
 measures 225
 patient safety 15, 20, 49, 599, 609,
 733, 763, 828, 973, 994, 1038,
 1048, 1066, 1070, 1075, 1080,
 1095, 1108, 1351, 1365, 1368,
 1369, 1379, 1388
 patient selection 280
 patient transfer 999
 patient-centered care 113, 1227
 patient-reported 103
 patients 253
 pediatric obesity 183
 pediatrics 275
 Pedigree 1278
 perceptions 689
 person-centered design 216
 personal 202, 221, 1234
 personal health record(s) 30, 79,
 216, 1228
 personal satisfaction 1311
 personally identifiable
 information 704
 pharmaceutical databases 920
 pharmaceutical services 1080
 pharmacoepidemiology 1251
 pharmacogenetics 1255
 pharmacovigilance 322, 614
 phenotype 599

phenotypic analysis 581, 1285
 physical activity 136
 physical therapy specialty 1321
 physician-patient relations 193
 physicians 103
 physiology 198
 point-of-care systems 1313, 1360
 point-of-care testing 471
 polysomnography 268
 population health 544
 practice guideline(s) 1220, 1336
 practice management 1158
 precision medicine 263, 366, 442,
 491, 564, 700, 1259, 1354
 pregnancy 221
 pregnancy complications 1336
 prenatal diagnosis 853
 prescriptions 594, 653
 pressure ulcer 1113
 primary care 103, 1342
 primary health care 35, 1230
 privacy 202, 704, 1312
 privacy preserving data 1303
 problem-oriented 1358
 process mining 554
 professional competence 679, 783
 prognosis 118, 1280, 1291
 program development 249
 program evaluation 1236, 1325,
 1335, 1364
 prostate cancer 1273
 psychiatric nursing 1235
 psychiatry 930
 psychological 98
 psychology 1302
 psychotherapy 193
 public health informatics 361, 693,
 696, 1282
 public health surveillance 808
 public health systems
 research 1335
 publications 437, 968
 PubMed 207, 873
 pulmonary embolism 1363
 pulse condition 586
 pulse wave analysis 573
 pupil 589

Q

quality and safety 955
 quality assurance 863, 1090, 1371
 quality improvement 1033
 quality indicators 1004, 1103
 quality of health care 1343
 quality of life 412

R

radiation dosage 1250
 radiography 1029, 1250
 radiology 1356

radiology information systems
 1325, 1364
 rare diseases 536, 896
 reading center 1254
 reference standards 30, 337, 868,
 887, 1327
 registries 303, 361, 536, 1143,
 1382
 regression analysis 573
 rehabilitation 74, 1217, 1374
 relational data management 950
 renal insufficiency 1232, 1354
 reports 1356
 research 1257
 resource allocation 1286
 respiratory insufficiency 417
 respiratory rate 417
 retina 289, 318
 review 1216, 1380
 rheumatoid 1223
 risk 639
 risk assessment 398, 578, 1158,
 1244
 risk factors 892, 1113
 risk management 609, 973, 1387
 robotics 1226
 RxNorm 843, 1333

S

safety management 1163, 1340,
 1385
 sample size 1279
 sarcoma 1375
 Saudi Arabia 1210
 school health services 1239
 screening behavioral risks 275
 security policy 1133
 sedentary lifestyle 240
 self care 1232, 1244
 self-management 131
 semantic annotation 581
 semantics 313, 437, 501, 657, 858,
 915, 945, 1285, 1328, 1341,
 1347, 1382
 sepsis 1367
 service optimization 1343
 sexually transmitted diseases 361
 Sjögren's syndrome 1166
 smartphone 10, 25, 156, 161, 268,
 1214, 1234, 1272
 smoking 1281
 smoking cessation 5
 SMS 1209
 SNOMED CT 848
 social determinants of health 486
 social factors 930
 social media 122, 322, 408, 1218,
 1219, 1297
 social networking 198, 240, 1229
 social support 1219

software 375, 1185, 1264, 1309, 1362
 software design 253, 1085
 software engineering 1262
 sound 526
 South Africa 1300
 squamous cell 634
 standardisation 1342
 standards 1371, 1378
 staphylococcal skin infections 1239
 statistical 662, 868, 915, 1323
 statistical data 662
 statistical models 332
 stochastic processes 332
 stress 98
 stroke 1024, 1195, 1216
 students 1243, 1284
 substance-related disorders 594, 1304
 suggestion 174
 supervised machine learning 298, 398, 604
 surgical wound infection 955
 surveys and questionnaires 753, 793, 858, 901, 1221
 survival analysis 1263
 system integration 1254
 systematic 1216
 systematized nomenclature of medicine (SNOMED) 925, 1332, 1334, 1341
 systemized nomenclature of medicine 1333
 systems integration 1275

T
 tailoring 379
 Taiwan 1353
 task performance and analysis 1080
 taxonomy 126
 technology 45, 818, 945, 1229, 1307
 technology acceptance model 818
 technology assessment 823
 telehealth 1223
 telemedicine 5, 20, 30, 69, 94, 258, 1063, 1212, 1213, 1220, 1221, 1231, 1233, 1234, 1239, 1240, 1253, 1313, 1320, 1329, 1339, 1343, 1360
 tere-care system 1366
 terminology 1332, 1337, 1345
 terminology as topic 940
 text mining 457, 955
 theoretical 84, 174, 506, 696, 778
 theoretical models 639
 thermography 384, 417
 three-dimensional 1237
 thyroid nodule 384
 topic coverage 978
 topical 920
 traditional 1257
 traditional Chinese medicine 586
 transportation 1383
 trauma and stressor related disorders 351
 triage 1214
 trust 202
 trusted health information 1004

type 1 619

U
 unified medical language system 501, 657, 925, 1252
 United States 1236
 universal coverage 738
 unsupervised machine learning 624
 user-computer interface 244, 763, 768, 1066, 1128, 1138, 1217

V
 vaccination 1218
 venous thromboembolism 1175
 veterans 1357
 virtual navigation 1375
 visually impaired persons 1381
 vital signs 1346
 vocabulary 235, 511, 843, 863, 887, 1327, 1332, 1334
 vulnerable populations 442

W
 waiting lists 989, 1286
 warfarin 370
 weight loss 379
 weights and measures 1195
 wireless technology 1387
 workflow 788, 835, 1270, 1368
 wound healing 1237
 WTP: (willingness to pay) 1366

This page intentionally left blank

Author Index

A

Aaron, S. 1385
 Abaza, H. 1209
 Aboukhamis, R. 322
 Abuzabaida, F. 1212
 Adam, T.J. 370, 614, 1128
 Adams, M.F. 1255
 Adlassnig, K. 1009, 1190
 Adlassnig, K.-P. 1009, 1123, 1190, 1336
 Affleck, E. 696
 Aguiar, A. 235
 Ahonen, O. 679
 Ai, A. 1385
 Aimé, X. 930
 Ainsworth, J. 45
 Ajay, V.S. 1230
 Ajmi, F. 989
 Akahori, S. 1279
 Aki, M. 1278
 Al Redha, M. 64
 Alassia, L. 709, 835
 AlBarrak, A. 1211
 Albergo, I. 1375
 Albiero Berni, J.C. 244
 Albornoz, M.A. 994
 Albuquerque, E.S. 198
 Alenazi, H. 1210
 Alghamdi, M. 1210
 Aljunid, S. 1334
 Alloni, A. 1374
 Al-Natour, A. 1212
 AlOthman, R. 1211
 Alradhi, S. 1210
 Al-Shorbaji, N. 1212
 Alvarez, G.G. 696
 Amatayakul, M. 308
 Ambastha, A.K. 1249
 Amith, M. 838
 Ammon, D. 1378
 Anand, A. 280
 Anand, V. 275, 280
 Andersen, J.M. 1385
 Andor, M. 526
 Antunes, K.T. 1250
 Aponte-Tiniao, L.A. 1375
 Appel, A.P. 412
 Appel, R.D. 1004
 Araki, K. 1305
 Aramaki, E. 285, 408
 Arno, J.N. 361
 Årsand, E. 619
 Arsoniadis, E.G. 955, 999

Ash, J.S. 700, 1385
 Ashraf, Md.M. 773
 Astudillo, J. 1319, 1346
 Aswath, A. 559
 Augustin, U. 788
 Avendaño, A. 693
 Ayerza, M.A. 1375
 Azuma, M. 1278

B

Babic, A. 240
 Bacigalupo, J.C. 1384
 Bada, M. 644
 Bagesteiro, L.D. 289
 Bahn, S. 1368
 Bahr, N.J. 843
 Bai, T. 873
 Bak, P. 332
 Balatsoukas, P. 79
 Ball, M.J. 98, 1004, 1351
 Ballout, G. 1212
 Bamidis, P.D. 5
 Barnett, J. 178
 Baron, S. 700
 Barreto, R. 559
 Bartfai, A. 1219
 Barton, D.L. 131
 Bastos, R.M. 1029
 Bates, D.W. 1251, 1385
 Batteiger, T.A. 361
 Baum, A. 709, 813, 835, 1213
 Bayat, S. 303
 Becker, M. 294
 Beeler, P.E. 1251
 Bejarano, B. 689
 Ben Said, M. 1133
 Benítez, S. 212, 689, 1095, 1213, 1283, 1358
 Bergmans, J.-P. 1004
 Bertelsen, P. 793
 Besting, A. 1378
 Bettencourt-Silva, J. 1331
 Beukenhorst, A.L. 10
 Bhupatiraju, R.T. 1252
 Bian, J. 1270
 Bianco, S. 1282
 Bielinski, S.J. 1255
 Billis, A. 5
 Biondich, P.G. 442
 Bjerke, T.M. 999
 Blacky, A. 1009, 1190
 Blaes, A.H. 599
 Blair, D. 1255

Blair, I. 64
 Blais, R.K. 351
 Blaser, J. 1251
 Blätzing, M. 1382
 Blobel, B. 202
 Boaretti dos Santos, F. 589
 Böckmann, B. 294, 1241
 Bodenreider, O. 662, 843, 853, 920, 925, 1252, 1333
 Boguslav, M. 298
 Boockvar, K.S. 1357
 Borda, A. 457
 Borges Tomaz, F. 1287, 1380
 Borges Vicente, R. 1380
 Born, J. 230
 Borthne, A. 1364
 Borycki, E.M. vii, 15, 178, 763, 1066, 1108, 1138
 Botsis, T. 619
 Bouaud, J. 1148
 Bouayad, L. 1261
 Bouillon, P. 644
 Bouix-Picasso, J. 1284
 Bourdin, E. 1099
 Bouslimi, D. 303
 Bouzillé, G. 303
 Bowen, M.E. 684
 Boyce, R.D. 960
 Boyd, A.D. 308
 Boyer, C. 1004
 Bozikov, J. 1320
 Bozkurt, S. 1325
 Bradshaw, D. 1300
 Brammen, D. 1315
 Brandt, C. 1261
 Bray, B.E. 327
 Brignone, E. 351
 Brixey, J. 1253
 Brixey, J.J. 1253, 1297
 Brochhausen, M. 960
 Browne, A.C. 501
 Bruchanski, L. 216, 1038
 Bruland, P. 313, 858, 1254
 Buchan, I.E. 79
 Buckeridge, D.L. 1335
 Burg, M.M. 1292
 Burgun, A. 322
 Bürkle, T. 20
 Burman, P. 506
 Burstein, G.R. 594
 Busnel, Y. 303
 Butler-Henderson, K. 1299

C

Caffarel, J. 554
 Caicedo, I.T. 1220
 Callahan, T.J. 644
 Camilo, E.N.R. 589
 Campos, F. 1213
 Cano, R. 212, 689
 Canteros, F. 1227
 Cao, Y. 1351
 Capurro, D. 693, 1377
 Caraballo, P.J. 1255
 Carpentier, M. 1004
 Carter, M.E. 351
 Carvalho, S.T. 198
 Cattani, B. 1374
 Cauvin, J.-M. 303
 Cerentini, A. 318
 Ceretta, L.B. 1287, 1380
 Cervi Prado, A.L. 74, 1217
 Cesconetto, S. 1287
 Chae, J. 1365
 Chai, R. 573
 Champion, C. 696
 Chan, T.-M. 398
 Chan, W.-M. 624
 Chang, C.-M. 808
 Chang, H. 98
 Chang, P.-L. 161, 268, 818,
 1014, 1214, 1232, 1233,
 1243, 1256, 1363
 Charlet, J. 930
 Chase, D.A. 700
 Chau, P.Y.-K. 624
 Chen, E.S. 486
 Chen, G.-T. 268
 Chen, H. 1163, 1352
 Chen, J.H. 768
 Chen, P.-T. 1256
 Chen, Q.-G. 1233
 Chen, W.-M. 1233
 Chen, X. 126, 322
 Chen, Yalan 263
 Chen, Yan 863, 1330
 Chen, Yong 1218, 1268
 Chen, You 1019
 Cheng, C.-L. 25
 Cheng, H.-W. 1256
 Cheng, P.-L. 818, 1214, 1363
 Cheng, Y. 327
 Cheng, Y.-T. 25
 Chi, C. 174, 511
 Chi, S. 1263
 Chiang, M.-C. 1215
 Chida, K. 1303
 Chien, C.-E. 1233
 Chihab, J. 1321
 Cho, I. 1258, 1365
 Cho, J. 467
 Cho, S.-Y. 467, 578

Choi, H.-J. 146, 1226, 1235,
 1274
 Choi, I.Y. 1273
 Choi, J. 452
 Choi, K.-S. 1298
 Choi, M.J. 1273
 Choi, S.Y. 1272
 Chorev, M. 332
 Chou, W.-J. 1353
 Chronaki, C. 30
 Chuang, J.-H. 808
 Chung, J. 1258
 Chute, C.G. 337, 940
 Cilliers, L. 35
 Cimino, J.J. 341
 Cintho, L.M.M. 1322
 Clark, S. 40
 Coatrieux, G. 303
 Cohen, A. 40
 Cohen, K.B. 298, 346, 644
 Cohen, T. 657
 Coiera, E. 609
 Comunello, E. 1287, 1380
 Conci, A. 384
 Constenla, D. 40
 Cordeiro d'Ornellas, M. 59,
 74, 244, 289, 318, 1029,
 1217, 1250, 1356
 Cornet, R. 848
 Cornford, T. 1080, 1259,
 1369
 Côté, J. 1284
 Covich-Bordenick, J. 1236
 Cronkite, R. 768
 Cucci, F. 1331
 Cuggia, M. 303
 Cui, L. 531, 1267
 Cui, T. 1352
 Cui, Y.-Y. 268, 1232, 1243
 Cullen, C. 1331
 Cullen, T.A. 442, 1354
 Cunningham, J. 45
 Cypko, M.A. 1323, 1355

D

da Costa, R.M. 589
 da Rosa Brunel Alves, P.A.
 1287
 da Silva Santos, B.H. 1322
 Dahamna, B. 322
 Dahm, M.R. 471
 Dalal, S. 1090
 Dalianis, H. 393
 Dalló, E.D. 1287
 Damião, C. 384
 Darlison, M.W. 1321
 Darmon, D. 462
 Das, S. 113
 Daus, M. 49, 221

Davidson, K.W. 1292
 Davis, R.L. 1335
 Day, K. 1239
 de Albuquerque, E.S. 589
 de Bruin, J.S. 1009, 1123,
 1190, 1336
 de Carvalho, D.R. 1380
 de Keizer, N. 848
 de los Hoyos, J. 1377
 de Melo Nunes Soares,
 F.A.A. 589
 de Moraes, J.P. 1217
 de Oliveira, L.L.G. 589
 de Santana, V.F. 412
 de Souza Pires, M.M. 1287,
 1380
 de Vries, G.-J. 554
 Defossez, G. 303
 Deldossi, M. 322
 Denecke, K. 20, 54, 1260
 Deng, Y. 1260
 Deparis, S. 1331
 Deserno, T.M. 1237, 1262
 Dey, A.B. 1307
 Dey, S. 113
 Dhombres, F. 662, 853
 Dias, K.N. 59
 Diaz, K.M. 1292
 Dickson, J. 1259
 Dieter, M.G. 308
 Dietz, A. 1355
 Dimaguila, G.L. 1216
 Ding, B.-F. 1014
 Dintelman, S. 103
 Divita, G. 351, 356
 Dixon, B.E. 361, 1357
 Dixon, W.G. 10
 Dong, J. 263
 Dong, W. 398
 Doods, J. 313, 858
 Dotto, G.N. 59, 289, 318,
 1029, 1250
 Dourado, E. 1227
 Dozo, M. 1227
 Du, J. 1218, 1268
 Du, X. 491, 1024
 Duan, H. 398, 427, 549
 Duan, H.M. 1293
 Duan, R. 1218, 1268
 Duboi, C.-A. 1284
 Dugas, M. 225, 230, 313, 858,
 1175, 1254
 Dutra da Costa, L. 1356

E

Echeverry, A. 693
 Edmilson da Silva Maia, C.
 244
 Eghdam, A. 1219

Eicher, J. 704
 Eklund, B. 798
 El Helou, S. 1266
 Elbert, T. 40
 Elers, P. 1229
 Elhanan, G. 978, 1330
 El-Hassan, O. 64
 Elison, J.T. 1128
 Elkin, P.L. 594, 1324
 Ellingsberg, A. 748
 Embi, P. 1385
 Erika, H. 1058
 Erlbaum, M.S. 920
 Escobar, G. 1220
 Esteban, S. 366, 709
 Eter, N. 1254

F

Fan, X.L. 604
 Fan, Y. 370
 Fang, Y. 1301
 Farfalli, G.L. 1375
 Fargo, J.D. 351
 Farnia, T. 69
 Faulbrück-Röhr, S. 1378
 Faulstich, L. 1260
 Faviez, C. 322
 Fensli, R. 1222
 Fernando, E. 1143
 Figueroa, R. 693
 Finch, D. 1261
 Flores, B. 375
 Flórez-Arango, J.F. 1220
 Fodeh, S.J. 1261
 Foletto, A.A. 74
 Fonck, S. 1262
 Fossler, R. 1190
 Fraccaro, P. 79
 Frangella, J. 216
 Franke, H. 1321
 Franklin, B.D. 1259
 Fu, G. 1341
 Fujii, A. 1372
 Fumihiro, H. 1366
 Fung, H. 624
 Fung, K.W. 1252, 1338
 Fung, V. 1342

G

Gabrielson, D.B. 1255
 Gaebel, J. 1323
 Gaff, C. 89
 Gagnon, M.-P. 1284
 Gaiera, A. 1358
 Galindo, C. 693
 Gallos, P. 1221
 Galopin, A. 1148
 Gamo, S. 1303
 Ganzinger, M. 375

Gao, C. 1019
 Gao, X. 714
 Gao, Y. 1263
 Gaouar, M. 1148
 García, A. 693
 Garcia, D. 1322
 Garcia, G. 216
 García, M.G. 1085
 García-Martínez, R. 1282
 Garthaus, M. 823
 Gartner, D. 84
 Garvin, J.H. 327
 Garza, T. 1255
 Gaudinat, A. 1004
 Gaudioso, C. 1324
 Gefeller, O. 901
 Geissbühler, A. 1004
 Geller, J. 863, 978
 Geng, Z. 156, 253
 George Jr., T.J. 1270
 Georgiou, A. 471, 569
 Gerdes, M. 1222
 Germano, E. 198
 Gerrett, D. 1369
 Gesicho, M.B. 718
 Geßner, S. 225, 858
 Gessner, S.I. 1175
 Ghosh, S. 280
 Ghres, R.A. 1250
 Gibbings, R.A. 1359
 Giraldo, L. 1095
 Girardeau, Y. 322
 Giri, J. 1255
 Giussi, M.V. 216, 709, 813, 835
 Goel, S. 935
 Gogia, S.B. 1264
 Goldschmidt, Y. 332
 Goldstein, M.K. 768
 Golse, B. 1133
 Gonçalves, C. 589
 Gonçalves, R. 235
 Gong, L. 873
 Gong, Q.-y. 586, 653, 1257
 Gong, Y. 684, 733, 828, 973, 1048, 1070, 1075
 González Bernaldo de Quirós, F. 709, 813, 835, 1085, 1213, 1375
 González, J.R. 384
 González, Z. 212, 1095
 Gorfer, T. 945
 Goss, F.R. 346
 Gottlieb, A. 768
 Gou, L. 1153, 1180
 Grain, H. 1228
 Grainger, R. 1223
 Grannis, S.J. 442
 Gray, K. 89, 457, 1216, 1299

Greenes, R. 758
 Greiner, L.E. 178
 Griffith, J. 94
 Grimsgaard, S. 108
 Grummet, J. 1360
 Gu, C. 491
 Gu, Y. 1224
 Guenegou-Arnoux, A. 322
 Guesgen, H. 1229
 Guillemin-Lanne, S. 322
 Guillen, S. 1213
 Guillermin, T. 103
 Guíñez, S. 693
 Gul, N. 1369
 Gulbrandsen, P. 1364
 Gunashekar, D.R. 614
 Gundlapalli, A.V. v, 351, 356
 Gunn, M.L. 1090
 Günther, A. 1276, 1337
 Guo, H. 388
 Guo, M. 915, 1362
 Guo, S. 1024
 Guo, Y. 1270
 Gupta, D. 183
 Gupta, P. 1230

H

Hababeh, M. 1212
 Haddad, P. 1311, 1313, 1360
 Häggglund, M. 723, 753
 Hahn, U. 521
 Hailey, D. 892, 1118, 1352
 Hakala, T. 1304
 Haku, I. 1103
 Håland, J. 1381
 Hall, C.S. 1090
 Halper, M. 978
 Hamidi, U. 1219
 Hammadi, S. 989
 Han, H. 1225
 Han, S.-H. 1226
 Han, S.J. 1272
 Han, W. 118
 Hanbury, A. 1004
 Hanlen, L. 1312
 Hanmer, L. 1300
 Hansen, D. 783
 Hao, L.-I. 573
 Hao, T. 98
 Harder, K.A. 1128
 Harris, M.R. 131, 838
 Härtel, S. 693
 Hartvigsen, G. 108, 619
 Hasegawa, S. 1303
 Hattori, A. 516
 Haugland, D. 748
 Hauskrecht, M. 1385
 Haux, R. 788, 901
 Hayen, A. 569

He, D. 263
 He, S. 103
 He, Y. 714, 1163, 1362
 He, Z. 863, 1265, 1270
 Healy, T. 170
 Heidenreich, P.A. 506
 Heitman, E. 718
 Helm, G. 1337
 Hemarathna, A. 1143
 Hennessy, C. 1331
 Henriksen, A. 108
 Henriksson, A. 393
 Herasevich, V. 1309
 Hercigonja Szekeres, M. 1320
 Herre, H. 1378
 Hertzum, M. 1033
 Hewapathirana, R.H. 1286
 Hibberd, R. 1259
 Hickerton, C. 89
 Hickman, T.T. 1385
 Hill, C. 848
 Hiltunen, R. 803
 Hines, L.C. 1255
 Hiragi, S. 1266
 Hirokawa, S. 649
 Hiroshi, K. 1058
 Hiroyuki, M. 1366
 Hisashi, E. 1366
 Ho, K. 249
 Hochstrasser, D. 1004
 Hofer-Dückelmann, C. 1123
 Holm, J. 20
 Honda, M. 1361, 1373
 Honey, M. 1239
 Hong, N. 868
 Hong, Y. 1325
 Hong, Y.A. 1301
 Hoover, S. 361
 Hopstock, L.A. 108
 Hörbst, A. 945
 Hose, D. 375
 Hou, C.-h. 1214
 Hou, I.-C. 1353, 1370
 Hou, Y. 1267
 Househ, M. 1210, 1213
 Hovenga, E. 1228
 Hsueh, P.-Y.S. 113
 Hu, D. 398
 Hu, G. 639, 1185, 1244
 Hu, H. 714, 915, 1362
 Hu, J. 491
 Hu, K.-f. 653
 Hu, M. 118, 403
 Hu, Z. 955
 Huang, B. 511
 Huang, C.Y. 1363
 Huang, J. 1218, 1268
 Huang, L. 873
 Huang, Z. 398
 Hübner, U. 1302

Hudson, E. 1284
 Huff, S.M. 337
 Hullin, C. 1213
 Hultman, G.M. 1128, 1269
 Hunter, I. 1229
 Hunter, L.E. 346, 644
 Huo, T. 1270
 Hurlen, P. 1364
 Huusko, J. 878
 Hwang, S.-h. 141
 Hyeon, J. 1235

I

Iles-Mann, J. 471
 Imai, T. 432, 882, 910
 Indarte, S. 1213
 Ingrand, P. 303
 Inoue, T. 1379
 Irino, S. 1303
 Iriso, S. 1240
 Ishida, H. 1303
 Islam, Md.M. 122
 Iso, H. 408
 Ito, M. 412
 Ivory, C. 1019
 Iwai, S. 882
 Iwasawa, M. 1340
 Izurieta, F. 1377

J

Jacob, P. 1230
 Jacquelinet, C. 303
 Jais, J.P. 1133
 Jang, C. 118
 Jara, D. 1377
 Jarabek, B. 999
 Jaulent, M.-C. 69, 930
 Jáuregui, O.I. 1038
 Jayatilleke, W.M.A.U. 1286
 Jenders, R.A. 1326
 Jeong, H. 417
 Jezek, E. 644
 Jha, D. 1230
 Jheng, B.-W. 1370
 Jheng, Y.-W. 1370
 Ji, X. 422
 Ji, Z. 126
 Jia, Z. 427, 549
 Jiang, G. 868, 887, 1327
 Jiang, K. 263
 Jiang, T. 892
 Jiang, Y. 131
 Jin, F. 1277
 Jin, S. 1271
 Jindal, D. 1230
 Johnson, T. 657, 662
 Jones, J.F. 728
 Joo, S.W. 1272
 Jortveit, J. 1222
 Judd, T. 447

Judkins, J. 960
 Juhn, Y.J. 1170
 Juhra, C. 230
 Julian, J. 1292
 Jung, H. 1043
 Jung, M. 1365
 Jung, M.-Y. 1367, 1368
 Jung, S. 467

K

Kadam, A. 559
 Kaeding, A. 1378
 Kagawa, R. 432
 Kahn Jr., C.E. 896, 1325
 Kalibatseva, Z. 1224
 Kamel Ghalibaf, A. 379
 Kang, G.-R. 136
 Kang, H. 733, 828, 973, 1048
 Karapetiantz, P. 322
 Karara, G. 738
 Kariyawasam, N. 743
 Karow, M.E. 1255
 Karvonen, T. 1266
 Kasáč, Z. 437
 Kasiiti, N. 1053
 Kasivajjala, V. 559
 Kasthurirathne, S.N. 442, 1354
 Kathöfer, U. 1254
 Katsahian, S. 322
 Katsuhiko, O. 1366
 Katta, S. 728
 Kautto, A.-K. 1304
 Kawazoe, Y. 432, 882
 Kazienko, J.F. 289
 Kei, T. 1058
 Kelly, K. 1331
 Kerns, R.D. 1261
 Khader, A. 1212
 Khairat, S. 599
 Kho, A.N. 935, 1019
 Khodambashi, S. 748
 Khumrin, P. 447
 Kiefer, R.C. 887, 1327
 Killgore, E.A. 684
 Kim, D. 1271
 Kim, D.-Y. 141
 Kim, E.Y. 1272
 Kim, H.-G. 1226, 1274
 Kim, I.K. 136, 141
 Kim, Ji. 1365
 Kim, Ju. 1231
 Kim, J.K. 1273
 Kim, K. 452
 Kim, M.-G. 141
 Kim, M.J. 1272
 Kim, S. 1334
 Kim, T.Y. 1271
 Kim, W.J. 136
 Kim, W.Y. 1367

Kim, Ye. 1365
 Kim, Yo. 1274
 Kim, Y.-J. 1367
 Kim, Z.M. 146
 Kimura, E. 1303
 Kinnunen, U.-M. 679
 Kiossoglou, P. 457
 Kishi, N. 1340
 Kitajima, A. 235
 Klecun, E. 1259
 Kleine, M. 1175
 Knaup, P. 375
 Ko, B. 1235
 Ko, C.Y. 955
 Ko, J. 1271
 Kobayashi, S. 1266, 1305,
 1308, 1328
 Koch, S. 753, 1219
 Koeda, N. 1379
 Koivunen, M. 1304
 Koller, W. 1009, 1190
 Komino, K. 1379
 Kondo, E. 1266
 Konstantinidis, E. 5
 Kopanitsa, G. 1063, 1306,
 1329, 1371
 Kopitowski, K. 366
 Kotoulas, S. 1331
 Kottum, H. 748
 Kozaki, K. 910
 Krebs, M.-O. 930
 Kreuzthaler, M. 539
 Krishnan, A. 1281, 1290
 Krishnan, C. 183
 Krudwig, S. 230
 Kuanghua, L. 1310
 Kuballa, S. 901
 Kuhn, K.A. 704
 Kulanthaivel, A. 728
 Kulikowski, C.A. 40, 758,
 873
 Kume, N. 1266, 1305, 1308,
 1328
 Kunikata, J. 1279
 Kuo, M.-C. 1014
 Kuo, T.B.J. 268
 Kurihara, Y. 1303
 Kuroda, T. 1266, 1308
 Kushniruk, A.W. 15, 178,
 763, 1066, 1108, 1378
 Kuusisto-Niemi, S. 878
 Kuziemy, C. 696
 Kvist, M. 393

L

Lacroix-Hugues, V. 462
 Lakshminarayana, R. 10
 Lancaster, W.J. 341
 Lander, S. 544
 Langford, A. 1265

Langlotz, T. 1223
 Lario, R. 920
 Lau, M. 1342
 Law, A. 634
 Le Guillou, C. 303
 Lee, D. 1235
 Lee, D.-Y. 136
 Lee, J.-H. 1231, 1367, 1368
 Lee, J.J. 1334
 Lee, J.S. 1273
 Lee, J.Y. 1273
 Lee, L.-H. 25
 Lee, Se. 467
 Lee, Su. 1271
 Lee, Sa.-W. 1367
 Lee, Si.-W. 141
 Lee, Y. 1231, 1368
 Lee, Y.-L. 1232
 Lehmann, M. 20
 Lei, J. 1243
 Lei, X. 915
 Leitersdorf, E. 768
 Leitch, H. 1336
 Lejonqvist, G.-B. 679
 Lemke, H.U. 1355
 Leonard, K.E. 594
 Leong, T.-Y. 1249
 Leslie, H. 905
 Leung, T.I. 768
 Li, H. 427, 549
 Li, H.-T. 1014
 Li, Jiao 388, 1288
 Li, Jing 1185
 Li, Ju. 1338
 Li, J.-s. 667, 1153, 1180,
 1263
 Li, J.-W. 268
 Li, L. 471
 Li, Xiang 476, 491, 639,
 1024, 1185, 1244, 1277
 Li, Xinhang 1263
 Li, Xiouchun 1236
 Li, Xu. 639, 1185, 1277
 Li, Y.-C. (Jack) 122
 Li, Z. 1274
 Liang, C. 973, 1070, 1075
 Liang, N.-W. 1233
 Liaw, S.-T. 773
 Lichtner, V. 1080, 1259, 1369
 Lillo-Le-Louët, A. 322
 Lim, C.-G. 146
 Lim, S. 1231
 Lin, C.-Y. 481
 Lin, K. 98
 Lin, S.-Y. 1353
 Lincoln, M.J. 920
 Lindberg, D. 1004
 Lindemann, E.A. 486, 1269
 Lindenfalk, B. 151
 Liou, D.-M. 481

Liu, C.-T. 1215
 Liu, D.-P. 808
 Liu, Haifeng 476, 491, 639,
 1024, 1185, 1244, 1277
 Liu, Hao 1330
 Liu, Ho. 581, 629, 1170, 1285
 Liu, H.-Y. 1233
 Liu, J. 118
 Liu, P. 1245
 Liu, S. 629
 Liu, Y. 156, 253
 Liu, Z. 667
 Ljosland Bakke, S. 905
 Lo, Y.-C. 1256
 Lockhart, C. 1229
 Lopez, V. 1331
 Lopez-Campos, G. 457, 496
 Lu, C.J. 501
 Lu, S.-C. 161
 Lu, X. 398, 549
 Lu, Y. 1263
 Luna, D.R. 49, 212, 216, 221,
 689, 994, 1038, 1085,
 1095, 1213, 1283, 1319,
 1346, 1358, 1384
 Luo, G. 356
 Luther, S. 1261
 Lv, X. 427
 Lyttkens, L. 798
 Lyttle, K. 578, 1158

M

Ma, J. 892
 Ma, P.-L. 1370
 Ma, Q. 1200
 Ma, X. 910, 1225
 Mabotuwana, T. 1090
 Machiraju, R. 422
 Maciel, A. 244
 Madden, M. 935
 Maeder, A.J. 166
 Magrabi, F. 609
 Mahajan, S.M. 506
 Mahumud, Y.S. 366
 Majeed, R.W. 1275, 1276,
 1289, 1315, 1337
 Makkad, P. 1166
 Maksim, T. 1306
 Malaviya, A.N. 1264
 Malin, B.A. 1019
 Mamlin, B.W. 442
 Manabe, S. 516, 1238, 1372
 Mandirola Brieux, H.-F. 1213
 Manenti, S.A. 1287
 Mangesius, P. 170
 Manion, F.J. 838
 Mann, D.K. 103
 Manohar, N. 486
 Mantas, J. 1221
 Mao, H. 174, 511

- Mao, W.-j. 653
 Marcelo, A. 773, 1213
 Marchand, G. 69
 Marco-Ruiz, L. 1332
 Markam, H. 778
 Marquard, J.L. 1128
 Marques da Silva, A.M. 244
 Márquez Fossier, S. 1358
 Márquez, S. 994
 Marschollek, M. 1209
 Martinez, R. 1384
 Martinez, S. 1222, 1381
 Martins, S. 1282
 Martin-Sanchez, F. 457, 496, 783
 Masaru, U. 1058
 Masud, J.H.B. 1213
 Matsamura, Y. 1372
 Matsumoto, K. 1280
 Matsumoto, T. 1361, 1373
 Matsumura, Y. 516, 1238
 Matsuura, Y. 417
 Matthies, F. 521
 Mayan, J. 1095
 Maydana, T. 49, 1095
 McBeth, J. 10
 Mccarthy, G. 1331
 McCaughey, E. 471
 McCreedy, L. 501
 McEvoy, D. 1385
 McEwan, R. 370, 1269
 McNicoll, I. 1321
 Meher, S.K. 1307
 Mehrotra, K. 1230
 Mei, J. 639, 1185, 1244, 1277
 Meidt, A. 858
 Meinel, C. 193
 Melton, G.B. 370, 486, 599, 955, 999, 1128, 1269
 Meng, H. 174, 511
 Merolli, M. 496, 1216
 Metcalfe, S. 89
 Miao, Q. 973, 1048
 Michael, E. 332
 Middleton, A. 89
 Mielke, C. 901
 Mihalas, G.I. 526, 758
 Mihara, N. 516, 1238, 1372
 Mihoko, O. 1103
 Milano, F.E. 1375
 Miller, R. 1004
 Mizushima, H. 536
 Modave, F. 1270
 Moen, A. 30
 Mole, E.S. 1283
 Monkman, H. 178
 Moon, T.D. 718
 Moore, J.H. 1268
 Moro, C.M.C. 1322
 Morris, K. 1308
 Moyano, L.G. 412
 Moyer, A.M. 1255
 Mugisho, E. 738
 Muguerza, P. 813
 Muhammad, I. 1313
 Mulero, R. 1240
 Müller, H. 1312
 Mulligan, N. 1331
 Mumford, V. 471
 Munyisia, E. 1118
 Murata, T. 516, 1058, 1372
 Murphy, T.F. 594
 Murthy, S. 1255
 Muscolo, D.L. 1375
 Musen, M.A. 768
 Mwogi, T. 1234

N
 Na, X. 388
 Nagasawa, T. 1278
 Nakagawa, A. 516, 1238, 1372
 Nakajima, T. 1278
 Nakamura, M. 403
 Nakashima, N. 403, 649, 1280
 Narasimhan, P. 773
 Narva, A.S. 1354
 Navas, H. 1099, 1227, 1282
 Ndabaniwe, E. 738
 Neira, R.A.Q. 554
 Nelson, S.D. 843, 920
 Neuhaus, P. 858
 Neumann, J. 1378
 Neumuth, T. 1378
 Newbold, S.K. 1297
 Newson, A. 89
 Newton, A. 506
 Nguyen, P.-A. 122
 Nguyen, T. 662, 1333
 Ni, Y. 604, 1244
 Nicholas, N. 569
 Nicola, E. 1300
 Nischwitz, A. 1382
 Nishimoto, N. 1279
 Nobuyuki, K. 1366
 Noedel, D.D. 1250
 Nohara, Y. 403, 1280
 Nolan, M.E. 1309
 Noppari, E. 1304
 Norton, J.M. 1354
 Nykänen, P. 803
 Nyong, J. 1230
 Nyssen, M. 738, 965
 Nytrø, Ø. 748

O
 Ochs, C. 978
 O'Connor, L. 1360
 Oeltze-Jafra, S. 1323, 1355
 Ogaito, T. 1387
 Ogasawara, K. 1345, 1383
 Ogata, H. 536
 Oger, E. 303
 Oh, E. 935
 Oh, K.-J. 1235
 Ohba, H. 1383
 Ohe, K. 432, 882, 910, 1376
 Ohno, Y. 417
 Ohno-Machado, L. 1312
 Okada, K. 1238
 Okamoto, K. 1266, 1308
 Okhmatovskaia, A. 1335
 Oleynik, M. 539
 Olsen, R. 1331
 Onimura, N. 649
 Oppermann, B. 788
 Ortiz, J.M. 1085
 Osmundson, S. 1019
 Otero, C.M. 1038, 1085, 1213, 1283, 1358
 Otero, P. 1213
 Othman, S.B. 989
 Overhage, J.M. 1200, 1236
 Overhage, L.M. 1236
 Oyama, K. 1225

P
 Padman, R. 84, 183, 672
 Pagidimarri, V. 559
 Pais, S. 188
 Pakhomov, S.V. 370, 1269
 Palermo, C. 709, 835
 Palojoki, S. 1108
 Pan, M.-L. 481, 808
 Panchatcharam, K. 183
 Panzarasa, S. 1374
 Paralescu, S. 526
 Paraskevopoulos, E. 5
 Park, C.-Y. 1226
 Park, H. 467
 Park, H.-A. 758, 1043, 1113
 Park, J. 1271
 Park, R.W. 467, 1271
 Park, S.K. 1113
 Park, T. 1368
 Park, Y.H. 1273
 Parker, J. 920
 Parker, J.C. 544, 578, 1158
 Parkulo, M.A. 1255
 Parry, D. 188, 983
 Parsons, F. 1292
 Patel, J. 1281, 1290
 Patel, V. 1230
 Pawola, L.M. 308
 Payne-Gagnon, J. 1284
 Pedersen, R. 1332
 Peek, N. vii, 79
 Peng, Y. 1310
 Peper, F. 366

Pereira Haygert, C.J. 59, 289,
318, 1029
Pereira, S. 1148
Perel, P. 1230
Perera, S. 1143
Perl, Y. 978, 1330
Perlich, A. 193
Peters, L. 1333
Peterschmitt, J.-C. 1004
Petersen, L.S. 793
Petrova, K. 188
Pettersson, M. 798
Phinney, M.A. 544
Pickering, B.W. 1309
Pistarini, C. 1374
Pladys, P. 303
Plasek, J.M. 968
Plazzotta, F. 813
Plössnig, M. 1123
Poly, T.N. 122
Popa, A. 788
Popejoy, L.L. 1158
Popescu, M. 578
Porthaine, F. 1378
Prabhakaran, D. 1230
Pradier, C. 462
Prasser, F. 704
Prodduturi, N. 868
Prosperi, M. 1270
Prud'hommeaux, E. 887,
1327
Przysucha, M. 1302
Purkayastha, S. 1053
Puttini, R. 235

Q

Qi, L. 573
Qian, S. 892, 1118
Qin, W. 427, 549
Quaglini, S. 1322, 1374
Quiros, F. 1283, 1319, 1346,
1358

R

Rajalahti, E. 679
Rajamani, S. 486
Raje, S. 925
Raju, M. 559
Ramaprasad, A. 564
Ramoni, R. 1385
Rao, P.S. 1255
Rapisarda, R.P. 1085
Rappelsberger, A. 1009,
1123, 1190, 1336
Ray, P. 773
Rebrij, R. 1384
Recondo, F. 835
Rector, A. 940
Redelico, F. 1319, 1346
Reid, D. 1118

Remmers, H. 823
Renard, J.-M. 989
Ribeiro, H.A. 198
Ricci, R.I. 366
Richard, M. 930
Rie, K. 1058
Riou, C. 303
Risk, M. 1085, 1319, 1346,
1375
Ritacco, L.E. 1375
Ritter, A. 422
Rizvi, R.F. 1128
Rizzato Ledo, D.A. 49, 1038,
1085
Robel, L. 1133
Robertson, H. 569
Rodrigues, J.-M. 940, 1334
Rodríguez Tablado, M. 366
Rodríguez, J.F. 1283
Röhrig, R. 1275, 1276, 1289,
1315
Rosenman, M. 935
Rouleau, G. 1284
Rowan, J. 188
Rowlands, D. 783
Roy, A. 1230
Rubin, L. 994, 1085, 1283,
1384
Rumana, U. 684
Runciman, W. 609
Rung, I. 1227
Ruotsalainen, P. 202
Ryan, A. 447
Ryan, P.B. 1200

S

Saadeh, R. 1212
Saboor, S. 170
Sacchi, L. 1322
Safran, C. 1004
Sakamoto, H. 1278
Sakurai, R. 910, 1376
Salanterä, S. 1312
Salazar, C. 1377
Saleh, K. 1378
Samaraweera, S. 1143
Samore, M.H. 351, 356
Samuel, H.W. 207
San Román, E. 1319, 1346
Sandmann, S. 1175
Sangster, J. 103
Santos Machado, A. 1217
Saranto, K. 679, 878, 1108
Sarda, G. 599
Sargeant, A. 471
Sastry, N.K.B. 564
Sato, Y. 536
Satoshi, K. 1366
Sauver, J.St. 1255
Savard, J. 89

Sbodio, M. 1331
Scandurra, I. 723, 798
Schabetsberger, T. 170
Schachner, B. 212, 1095
Schaffer, J.L. 1311
Schaper, L. 783
Schmidt, T. 1378
Schneider, J. 960
Schreiber, R. 1385
Schreier, G. 1312
Schuler, J. 1123
Schultz, C. 788
Schultz, D.M. 10
Schulz, S. 437, 539, 940
Schulze, M. 901, 1209
Schwartz, J.E. 1292
Schweitzer, M. 945
Seckinger, A. 375
Seita, A. 1212
Senathirajah, Y. 15, 763,
1066, 1138
Seneviratne, K. 1143
Sergeant, J.C. 10
Séroussi, B. 1148
Seto, R. 1379
Seywerd, M.A. 1128
Shaban-Nejad, A. 1335
Shah, R.U. 327
Shahin, Y. 1212
Shamoon, F. 1336
Shang, Y. 263, 1153, 1180
Shao, Y. 327
Sharif, A. 64
Sharma, D.K. 887
Sheets, D. 178
Sheets, L. 544, 578, 1158
Shen, B. 1163
Shen, F. 581, 1285
Shen, M. 253
Shen, N. 253
Shi, C. 586, 1257
Shi, L. 263
Shi, Y.L. 604
Shimai, Y. 1372
Shin, E.K. 1335
Shin, G.W. 1368
Shinohara, E. 432, 910
Shishtawi, A. 1212
Shpigelman, L. 332
Shyu, C.-R. 544
Si, Y. 950
Siddiqui, Z. 1166, 1281, 1290
Silva, C.R.G. 589
Silva, N. 1339
Silva, S.N. 1286
Simões Pires, P.D. 1380
Simões, P.W. 1287, 1380
Simon, G.J. 955
Simon, M. 216
Singh, H. 1255

Sinha, A. 40
 Sinha, S. 594
 Siqueira, J. 589
 Sirazitdinova, E. 1237
 Sittig, D.F. 1385
 Skube, S.J. 599, 955, 999, 1269
 Slee, A. 1369
 Smaradottir, B. 1222, 1381
 Smith, M.I. 216, 221
 Soejima, H. 649
 Sogo, T. 1279
 Sohn, S. 1170
 Solbrig, H.R. 887, 1327
 Sommer, J.A. 216, 221
 Song, J.-H. 141
 Song, X. 1224
 Soto-Rey, I. 225, 858
 Spachos, D. 5
 Spalding, S.J. 275
 Specht, M. 1378
 Sri Prakash, B. 183
 Sriram Iyengar, M. 1220
 Staccini, P. 462
 Staemmler, M. 1382
 Stegwee, R. 30
 Stephen, R. 89
 Stieben, A. 835
 Stindel, E. 303
 Stoehr, M. 1355
 Stöhr, M.R. 1275, 1276, 1337
 Storck, M. 230, 313, 858
 Stretton, E. 554
 Stucke, S. 1378
 Sturm, J. 1382
 Su, Y.-C. 1214
 Sugamori, Y. 536
 Sugano, K. 1278
 Sui, M. 531
 Sun, D. 599
 Sun, H. 1338
 Sun, L. 1314
 Sun, Y. 1288
 Sunao, W. 1103
 Sund, J.K. 748
 Sundquist, K. 1292
 Suo, Y. 351
 Suomi, R. 803
 Suominen, H. 1312
 Sutton, J.A. 1255
 Syed Abdul, S. 122
 Syn, T. 564

T

Tahar, K. 1378
 Takahashi, D. 516
 Takeda, H. 1388
 Takeda, T. 516, 1238, 1372
 Tamborindéguy, A. 1384
 Tamura, H. 1266, 1308

Tamura, N. 1279
 Tana, J. 679
 Tanabe, M. 536
 Tanabe, N. 1278
 Tandon, N. 1230
 Tang, G. 1310
 Tang, G.Y. 604, 1244
 Tani, Y. 1279
 Tanikawa, T. 1345, 1383
 Tao, C. 838, 1218, 1268
 Tao, G. 361
 Tara, S.M. 379
 Taramasco, C. 693
 Taranik, M. 1371
 Taylor, W. 1223
 Teixeira, V. 1384
 Ten Napel, H. 1334
 Teppei, S. 1366
 Terrasa, S. 366
 Terrill, B. 89
 Texier, N. 322
 The Alzheimer's Disease
 Neuroimaging Initiative
 1249
 Thiemann, V.S. 1276, 1289, 1315
 Thonnet, M. 1004
 Thornton, S.N. 103
 Thyvalikakath, T. 1166, 1281, 1290
 Tian, Y. 1180, 1263
 Tieder, J. 1090
 Tissera, S. 1339
 To, D. 1342
 Toffanello, A. 235
 Tomita, M. 1340
 Topaz, M. 968
 Tormey, D. 501
 Touray, M. 122
 Townsley, H. 1223
 Tran, L.-T.T. 356
 Travaglia, J. 569
 Traversaro, F. 1319, 1346
 Treder, M. 1254
 Trombert, B. 1334
 Tsai, P.-Y. 1353
 Tsao, H.-M. 808
 Tsui, K.-L. 624
 Tsuyoshi, O. 1058
 Tu, W. 361
 Tudor, A. 526
 Tufte, T. 240
 Turbitt, E. 89
 Turki, Y. 1212
 Turner, P. 743
 Tye, M.L. 1239
 Tytherleigh, R. 89

U

Uciteli, A. 1378

Ueda, K. 1238, 1372
 Ushima, M. 1278
 Utecht, J. 960

V

van Bemmelen, J.H. 1004
 van der Veer, S.N. 79
 Van Laere, S. 965
 Varghese, J. 858, 1175
 Venson, J.E. 244
 Venters, W. 1259
 Vepsäläinen, T. 803
 Verbeke, F. 738, 965
 Verspoor, K. 447, 457
 Vidal, D. 1240
 Vidal, M. 689
 Vielhauer, J. 1241
 Vigo, M. 79
 Villalba, C. 1213
 Vimarlund, V. 151, 258
 Vogel, A. 230
 von Thienen, J. 193

W

Wakamiya, S. 285, 408
 Wallace, B.C. 657
 Wang, C. 868
 Wang, D. 1242
 Wang, D.-W. 808
 Wang, F. 1048
 Wang, Jane 361
 Wang, Jie 263
 Wang, Jinglu 1138
 Wang, Jingting 253
 Wang, J.L. 249
 Wang, Lanlan 1314
 Wang, Lei 1338
 Wang, Liwei 581, 1285
 Wang, Liqin 968
 Wang, P. 1014
 Wang, S. 118
 Wang, X. 1344, 1347
 Wang, Yan 486
 Wang, Yanling 1314
 Wang, Yanshan 629
 Wang, Ye 873
 Wang, Yefeng 614
 Wang, Ying 609
 Wang, Yu 1153, 1180
 Wang, Yuanyuan 253
 Wang, Yue 1166, 1290
 Wang, Y.B. 1293
 Wass, S. 258
 Wawira, J. 1053
 Weber, T. 230
 Wei, D. (Helen) 1341
 Wei, Y. 263
 Weir, C.R. 327
 Welfer, D. 59, 289, 318, 1029

Wen, A. 629
 Wen, T.-N. 161, 818
 Weng, C. 950
 Were, M.C. 718, 1053, 1234
 West, B.T. 131
 Westbrook, J.I. 471
 Westerlynck, R. 303
 Wetter, T. 113
 Whiddett, D. 1229
 Whipple, M.E. 634
 Wi, C.-I. 1170
 Wick, E.C. 955
 Wickramasinghe, N. 1311,
 1313, 1359, 1360
 Williams, P.A.H. 166
 Wilson, B. 89
 Winnenburgh, R. 843, 920
 Wohlmann, J. 230
 Woldaregay, A.Z. 619
 Wong, A. 1342
 Wong, E.L.-Y. 624
 Wong, M.C. 743
 Wong, Z.S.-Y. 624
 Workman, T.E. 356
 Wright, A. 1385
 Wright, G. 35
 Wu, B.-G. 1233
 Wu, C. 1153, 1180, 1386
 Wu, F. 156
 Wu, H. 263
 Wu, P. 268
 Wu, S. 629, 634
 Wu, X. 1386
 Wu, Y. 126, 1314, 1338
 Wüller, H. 823
 Wyatt, M.C. 341

X

Xavier, E.P. 1287
 Xia, E. 639, 1185
 Xiao, Q. 1245, 1314
 Xie, G. 476, 491, 604, 639,
 1024, 1185, 1244, 1277
 Xin, S. 1291
 Xu, E. 639, 1185
 Xu, H. 126, 838
 Xu, J. 126
 Xu, L. 573, 586
 Xu, M. 476, 639, 1185, 1277
 Xu, T. 1275, 1289, 1315
 Xu, Wei 1344, 1347
 Xu, Weiming 1244

Y

Yacubsohn, V. 758
 Yadav, N. 1331
 Yadav, P. 644
 Yaeli, A. 332
 Yagahara, A. 1345, 1383
 Yamada, M. 1278
 Yamamoto, G. 1266, 1308
 Yamamoto, Y. 516
 Yamashita, S. 1379
 Yamashita, T. 649
 Yamashita, Y. 1387
 Yan, L. 1310
 Yan, P. 1243
 Yang, H.-C. 122
 Yang, J. 476
 Yang, L. 1268
 Yang, Y. 476
 Yano, K. 285
 Yao, B. 973
 Yao, N. (Aaron) 253
 Yao, Y. 573
 Yasini, M. 69
 Yasoshima, T. 1379
 Yeh, Y.-T. 1215
 Yen, P.-Y. 422
 Yin, B. 1291
 Yin, H. 1243
 Yokoi, H. 1279, 1345
 Yook, I.H. 1273
 Yoon, D. 467, 1271
 Yoon, S. 1292
 Yoshida, T. 1278
 Yoshihara, H. 1305, 1328
 Yoshimune, S. 1103
 You, S.C. 467
 Young, C.D. 308
 Yu, F. 1362
 Yu, J. 511
 Yu, P. 892, 1118, 1352
 Yu, X.-w. 653
 Yu, Y. 1244
 Yu, Z. 657, 662
 Yuan, C. 156, 253
 Yuji, T. 1366
 Yuko, Y. 1058
 Yusuke, E. 1058

Z

Zaiane, O.R. 207
 Zakaria, N. 1210, 1211
 Zeckl, J. 1009, 1190

Zeft, A.S. 275
 Zeidan, W. 1212
 Zelechower, J. 1319, 1346
 Zeng, X. 549
 Zeng-Treitler, Q. 327
 Zetz, E. 20
 Zgaya, H. 989
 Zhang, B. 1245
 Zhang, E. 728
 Zhang, Han 1267
 Zhang, Hao 1267
 Zhang, J. 1325
 Zhang, L. 1277
 Zhang, P. 491
 Zhang, Qiannan 1314
 Zhang, Qiyao 1388
 Zhang, R. 370, 599, 614
 Zhang, W. 263, 1293
 Zhang, W.-m. 653
 Zhang, Xiaomei 1195
 Zhang, Xiaoyan 253
 Zhang, Xiaoying 1343
 Zhang, Xinyuan 1268
 Zhang, X.Z. 1293
 Zhang, Yaoyun 126, 838
 Zhang, Ying 667
 Zhang, Yiye 672
 Zhang, Yu 586, 1288
 Zhang, Yue 1362
 Zhang, Z. 361
 Zhao, D. 174, 511
 Zhao, S. 639, 1277
 Zhao, Y. 1291
 Zheng, L. 978, 1330
 Zheng, X. 549
 Zhou, H. 1195
 Zhou, J. 586, 1257
 Zhou, L. 968
 Zhou, Sh. 263
 Zhou, Si. 828, 973, 1048
 Zhou, T. 1153, 1263
 Zhou, Z. 1301
 Zhu, V.J. 1200
 Zhu, X. 98
 Zhu, Yafang 1195
 Zhu, Yanxin 1344, 1347
 Zhu, Ying 1325
 Zhuang, Y. 544
 Zippel-Schultz, B. 788
 Zolhavarieh, S. 983
 Zweigenbaum, P. 346

This page intentionally left blank